

AP 3302
Part 3
(2nd edition)

ELECTRONIC ENGINEERING

TRADE GROUP
(FITTERS)

RADAR



STANDARD TECHNICAL TRAINING NOTES
ELECTRONIC ENGINEERING TRADE GROUP

RADAR

By Command of the Defence Council



SI UNITS AND BS SYMBOLS

SI UNITS

Introduction

The intention to change gradually from the present system of units used in this country to the international *metric* system has been announced by the Government. Thus, more and more we shall begin to see millimetres being used instead of inches, kilogrammes instead of pounds, litres instead of gallons, and so on.

This move to change to the metric system coincides with the introduction of a new rationalized set of international units known as the *Système International* (SI). The SI system of units is, in the main, already familiar to us but it does involve a change of name of some of the units. Therefore it is important that we learn these units so that we may be able to recognise them and understand what is meant when we meet them.

SI Units

In any system of units, magnitudes of some physical quantities must be arbitrarily selected and declared to have unit value. These magnitudes form a set of standards and are called *basic units*. All other units are *derived units*, related to the basic units by definitions.

The SI system has six basic units. These, together with their symbols, are listed in Table 1. All other SI units are derived from these basic units.

Since many of the derived units in which we are interested are defined in terms of the *ampere*, it may be helpful to give the definition of the ampere:

Quantity	Name of Unit	Symbol
length	metre	m
mass	kilogramme	kg
time	second	s
electric current	ampere	A
thermodynamic temperature	degree Kelvin	°K
luminous intensity	candela	cd

TABLE 1. BASIC SI UNITS

'The unit of electric current called the ampere is that constant current which, if maintained in two parallel rectilinear conductors, of infinite length, of negligible cross section, and placed at a distance of one metre apart in a vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre length'. (See Table 2 for definition of newton.)

The other basic units also have precise definitions.

Special names and special unit symbols have been adopted for some of the *derived* SI units, as shown in Table 2 overleaf. For these units, the definitions show the relationships between them and the basic units, and the definitions are also included in Table 2.

Most of the units and their definitions given in Table 2 are familiar to us, but we should note particularly that:

- a. The unit for frequency is changed from *cycles per second* to *hertz* (Hz).
- b. The unit for magnetic flux density is the *tesla* (T).

Table 2 shows only the more common derived SI units. The SI units that are not listed can only be expressed in terms of the units from which they are derived, e.g. volume (cubic metre—m³) and acceleration (metre per second squared—m/s²).

Quantity	Unit	Symbol	Derivation	Definition
force	newton	N	kg m/s ²	One newton is that force which, when applied to a body having a mass of one kilogramme, gives it an acceleration of one metre per second squared.
work or energy	joule	J	Nm	One joule is the work done when the point of application of a force of one newton is displaced through a distance of one metre in the direction of the force.
power	watt	W	J/s	One watt is equal to one joule per second.
electric charge	coulomb	C	As	One coulomb is the quantity of electricity transported in one second by a current of one ampere.
electric potential	volt	V	W/A	One volt is the difference of potential between two points of a conducting wire carrying a constant current of one ampere when the power dissipated between these points is equal to one watt.
electric capacitance	farad	F	As/V	One farad is the capacitance of a capacitor between the plates of which there appears a difference of potential of one volt when it is charged by a quantity of electricity equal to one coulomb.
electric resistance	ohm	Ω	V/A	One ohm is the resistance between two points of a conductor when a constant difference of potential of one volt, applied between these two points, produces in this conductor a current of one ampere, this conductor not being the source of any electromotive force.
frequency	hertz	Hz	s ⁻¹	One hertz is the frequency of a periodic phenomenon of which the periodic time is one second.
magnetic flux	weber	Wb	Vs	One weber is the flux which, linking a circuit of one turn, produces in it an electromotive force of one volt as it is reduced to zero at a uniform rate in one second.
magnetic flux density	tesla	T	Wb/m ²	One tesla is the density of one weber of magnetic flux per square metre.
inductance	henry	H	Vs/A	One henry is the inductance of a closed circuit in which an electromotive force of one volt is produced when the electric current in the circuit varies uniformly at the rate of one ampere per second.

TABLE 2. SOME DERIVED SI UNITS AND THEIR DEFINITIONS

A few points are worth noting in connection with SI units:

- Names of units, even when commemorating a person, are not written with a capital initial letter—*eg* ampere, henry.
- The symbol for a unit named after a person has a capital initial letter—*eg* W (watt), Hz (hertz); but not m (metre) or s (second).
- Symbols for units do not have a plural form, with added 's'—*eg* 15m (not 15ms), 5kg (not 5kgs).
- The use of prefixes (see below) representing 10 raised to a power which is a multiple of 3 is especially recommended. From this it may be seen that the use of centimetres (10^{-2} m) is not recommended; millimetres (10^{-3} m) should be used instead.

Prefixes

Prefixes, by means of which the names of multiples and submultiples of units are formed, are shown in Table 3. The more common ones used in radio and electronic engineering are printed in **bold**.

Multiplication factor	Prefix	Symbol	Multiplication factor	Prefix	Symbol
1 000 000 000 000 = 10^{12}	tera	T	0.01 = 10^{-2}	centi	c
1 000 000 000 = 10^9	giga	G	0.001 = 10^{-3}	milli	m
1 000 000 = 10^6	mega	M	0.000 001 = 10^{-6}	micro	μ
1 000 = 10^3	kilo	k	0.000 000 001 = 10^{-9}	nano	n
100 = 10^2	hecto	h	0.000 000 000 001 = 10^{-12}	pico	p
10 = 10^1	deca	da	0.000 000 000 000 001 = 10^{-15}	femto	f
0.1 = 10^{-1}	deci	d	0.000 000 000 000 000 001 = 10^{-18}	atto	a

TABLE 3. MULTIPLES AND SUBMULTIPLES OF UNITS

Note from Table 3 that the comma is not used to separate groups of three digits in numbers containing more than four digits. Thus, we write 1345 or 0.0942, with no comma or space; but we write 12 351 or 0.321 45 with a space.

One of the most noticeable implications of the introduction of SI units is the adoption of the hertz (Hz) for frequency. We shall find the hertz, and its multiples (kHz, MHz, GHz), used more and more as time goes on. It is so used in the later stages of this book and for all pages that have been subject to amendment.

BS SYMBOLS

All graphical symbols used for electrical power, telecommunications and electronic diagrams should now conform with those listed in the British Standard BS3939, as amended.

Unfortunately, at the time this book was being prepared, BS3939 had not been issued. Therefore, all drawings in this book, up to and including AL10, use the pre-BS3939 symbols. The first use of the new symbols was in AL11, issued in October 1969. All subsequent corrections to this book will automatically include the BS3939 symbols.

Some commonly-used symbols are illustrated in Table 4 overleaf.

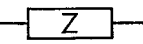
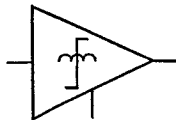
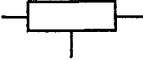
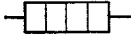
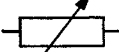
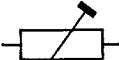
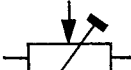
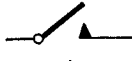
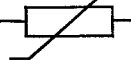
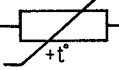
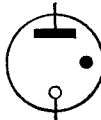
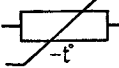
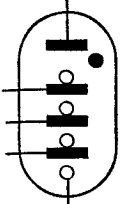
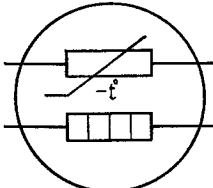
Description	Symbol	Description	Symbol
Impedance		Transformer	
Fuse		Transducer element in assembled representation	
Fixed resistor		Magnetic amplifier, general symbol for block diagrams	
Fixed resistor with fixed tapping (voltage divider)		Microphone	
Heater		Capacitor microphone	
Variable resistor		Earphone (receiver)	
Resistor with preset adjustment*		Headgear receiver, double (headphones)	
Resistor with moving contact		Relay coil	
Voltage divider with preset adjustment		Relay make contact-unit	
Resistor with inherent non-linear variability		Relay change-over (both sides stable) contact unit	
Resistor with pronounced positive resistance temperature coefficient, eg ballast resistor or barretter		Cold-cathode discharge tube, asymmetrical	
Resistor with pronounced negative resistance temperature coefficient, eg thermistor		Multi-electrode voltage stabilizer	
Indirectly-heated thermistor			
Winding (ie of an inductor, coil, choke coil, inductive reactor or transformer)			
If it is desired to indicate that the winding has a core it may be shown thus: Example: Inductor with core (ferromagnetic unless otherwise indicated)			

TABLE 4 EXAMPLES OF BS3939 SYMBOLS

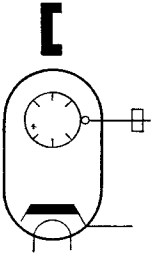
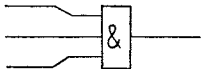
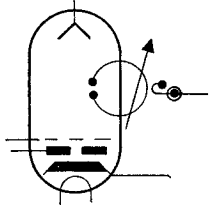
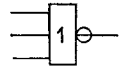
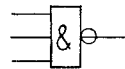
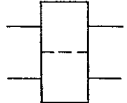
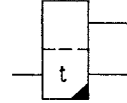
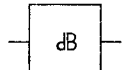
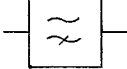
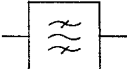
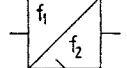
Description	Symbol	Description	Symbol
Magnetron oscillator tube with indirectly heated cathode, closed slow-wave structure with d.c. connection via waveguide, permanent field magnet and window-coupler to rectangular waveguide		AND element The output signal represents 1 when all the input signals represent 1.	
Reflex klystron with indirectly heated cathode, focusing electrode with aperture, grid, tunable integral cavity resonator, reflector and loop coupler to coaxial output		NOR element (NOT-OR element). The output signal represents 0 when one or more of the input signals represents 1. OR element. The output signal represents 1 when one or more of the input signals represents 1.	
pn diode		NOT-AND element (NAND element). The output signal represents 0 when all the input signals represent 1.	
pn diode used as capacitive device (varactor)		Bistable element with normal inputs and outputs. NOTE: A bistable element is the simplest example of a multistate element.	
Tunnel diode		Monostable element	
Unidirectional breakdown diode, voltage reference diode (voltage-regulator diode) eg Zener diode		Attenuator, fixed loss (pad)	
Thyristor		High-pass filter	
p n p transistor		Band-pass filter	
Unijunction transistor with p-type base		Frequency changer changing from f_1 to f_2	
P-channel insulated-gate field-effect transistor		Amplifier	

TABLE 4 (cont'd)

(AL12, November 1971)

CONTENTS

Section	Page
Foreword	
1. Introduction to Radar	1
2. Radar Circuits	61
3. UHF Radar	227
4. Centimetric Radar	249
5. Pulsed Radar Transmitters	361
6. Radar Receivers	395
7. CW Radar	445
8. Radar Data Handling	515
9. Some Modern Developments in Radar	543
Index	643

FOREWORD

1. This Air Publication has been prepared by Headquarters Training Command (Ed 1c) under the direction of the Ministry of Defence, Directorate of Training (Ground) (Royal Air Force).

2. It is published to assist those under training as fitters in the Electronic Engineering Trade Group. Fitters in this trade group require a thorough knowledge of the electrical, instrument, and radio principles appropriate to the theory of the specified equipment. It is with the intention of helping to attain this standard that this publication is written. It is not intended to form a complete textbook, but is to be used as required in conjunction with lessons and demonstrations given at the training school. It may also be used to assist airmen on continuation training at other RAF stations.

3. The AP 3302 series of Air Publications is sub-divided as follows:

Part 1A: Electrical and Radio Fundamentals. This deals with the principles of electricity, electronics, and radio at a level suitable for electronic technicians. Because of its bulk, Part 1A has been split into three separate books: Book 1, basic electricity; Book 2, basic electronics; and Book 3, basic radio.

Part 1B: Basic Electricity and Radio. This deals with the principles of electricity, electronics, and radio at a level suitable for electronic fitters.

Part 2: Communications. This deals with the applications of the principles covered in Parts 1A and 1B to communication systems and is intended to be used as required by electronic technicians and fitters.

Part 3: Radar. This deals with the applications of the principles covered in Parts 1A and 1B to radar and is intended to be used as required by electronic technicians and fitters.

Part 4: Navigation Instruments. This deals with the applications of the principles covered in Parts 1A and 1B to aircraft navigation instruments and is intended to be used as required by electronic technicians and fitters.

4. In general, fitters employed on communications equipment will be interested mainly in Part 1B and Part 2 of these notes; radar fitters will be concerned mainly with Part 1B and Part 3; and fitters (nav. inst.) will be interested mainly in Part 1B and Part 4. However, it is difficult to draw a firm dividing line between the knowledge required by electronic fitters in the various fields. There is considerable overlapping; for example, much of what was once regarded as being exclusively in the province of the radar fitter is now a requirement for the communications fitter also, and *vice versa*. Thus those under training in the radar field may find much that is useful in Parts 2 and 4 as well as in Part 3; communications fitters may find much of interest in Part 3 as well as in Part 2; and fitters (nav. inst.) may require information on radar navigational systems as given in Part 3, in addition to the instrument systems covered in Part 4.

5. The notes deal with the basic theory and the applied principles of the various subjects in a general way. They do *not* cover specific details of equipment in use in the RAF. Such details are to be found in the official Air Publication for the equipment and this should always be consulted during the servicing of the equipment.

6. The subject matter in this publication may be affected by the contents of Defence Council Instructions and by amendments to other official publications. Where this occurs, suitable amendment lists will be issued as soon as practicable. In the meantime, any mandatory document that contradicts information contained in this publication is to be taken as the over-riding authority.

(AL 12, Nov 71)

7. The material used in this publication has been gathered from numerous sources, both Service and civilian. An acknowledgement is extended to all who have contributed to its presentation.
8. Unauthorized alterations are not to be made to this book and, since some of the material may be subject to copyright, no reproduction of its contents, in whole or in part, is to be made without the prior permission of Headquarters Training Command.
9. Readers of this publication are asked to report any unsatisfactory features that they may notice. Such unsatisfactory features may include omissions, factual errors in words or illustrations, and ambiguous, obscure, or conflicting instructions. All such unsatisfactory features should be reported to Headquarters Training Command (Ed 1c) on RAF Form 6734.

Ministry of Defence (Air)
November, 1971

SECTION 1
INTRODUCTION TO RADAR

Chapter		Page
1	Pulse-modulated Radar	1
2	Basic Requirements of a Pulse-modulated Radar System	15
3	Factors Affecting the Performance of Pulse-modulated Radars	27
4	Some Examples of the Uses of Pulsed Radar	43
5	Basic Outline of CW Radar	53

CHAPTER 1

PULSE-MODULATED RADAR

Introduction

Radar is a method of using radio waves to detect the existence of an object and then to find its position in relation to a known point, usually the site of the radar installation (Fig 1). By means of radar the presence of moving or stationary objects such as aircraft, ships, and land masses can

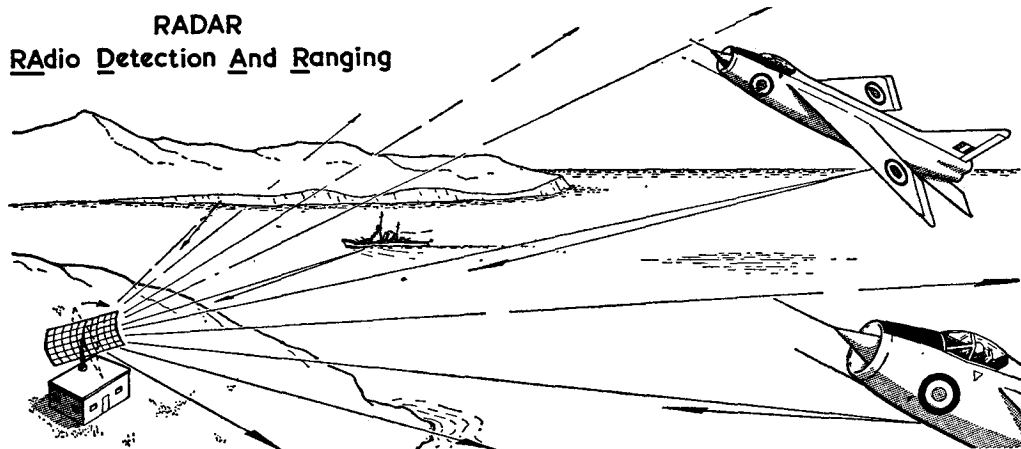


FIG 1. RADAR

be *detected*. In addition, information concerning the exact *position* of the object (usually referred to as the 'target') and its speed and course, where applicable, can be obtained.

The word '*radar*' is coined from the initial letters of the phrase:

RAdio Detection And Ranging.

Although radar was originally introduced to give warning of the approach of hostile aircraft it has since been further developed to do much more than its original task. Modern radar equipment plays a vital part in all the operational roles of the RAF: as an aid to accurate bombing; in airborne detection and interception equipment; for the control of guided weapons; as navigational and landing aids; in cloud and collision warning devices. It has many other uses in the civilian as well as in the Service field.

Pulse-modulated and CW Radar Systems

Most radar equipments are *pulse-modulated*, i.e. the radiation from the transmitter aerial is in the form of very short bursts or *pulses* of r.f. energy, each pulse being followed by a *relatively long resting period* during which the transmitter is switched off and the receiver is operating.

For certain applications pulse-modulated radar has limitations. A form of *continuous wave* (c.w.) radar is then used. This may be:

- a. Pure c.w. radar relying on the '*Doppler shift*' in frequency to detect moving objects and to measure their speed.
- b. *Frequency-modulated* c.w. radar where the difference in frequency between the reflected wave from a target and the direct wave from transmitter to receiver gives an indication of the range of the target.

Most of this book deals with pulse-modulated radar systems. However, a basic outline of c.w. radar is given in Chapter 5 and further details are given in Section 7.

Primary and Secondary Radar

Pulse-modulated radar systems can be divided into two main groups (Fig 2):

a. Primary radar. This relies on the *reflection* of a portion of the incident energy by the target. The target *does not willingly co-operate* with the radar installation. This is the type of radar used to detect and track hostile aircraft and missiles.

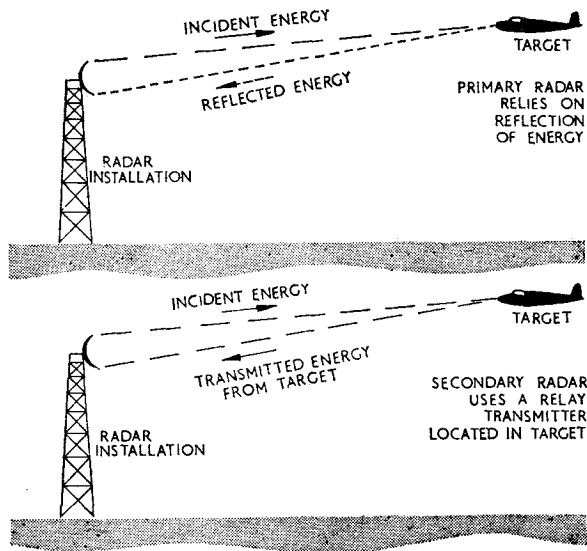


FIG 2. PRIMARY AND SECONDARY RADAR

b. Secondary radar. In this system the radar installation and the target *help each other*. The signal received at the radar installation is *not a reflection* of the incident energy but one from a *transmitter*, located in the target, which is switched on by the incident energy. An example of secondary radar is the identification equipment carried by all friendly ships and aircraft.

Reflection From Aircraft

When watching television you will probably have noticed that when a low-flying aircraft passes overhead the brightness of the screen increases and decreases causing the picture to 'flutter'. The reason for this is as follows:

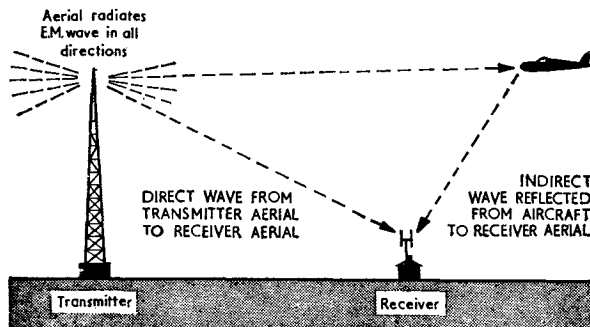


FIG 3. REFLECTION FROM AN AIRCRAFT

The transmitter aerial radiates an electromagnetic (e.m.) wave in all directions. The receiver aerial normally picks up only that part of the wave which travels in a *direct line* between the aerials. However, if part of the transmitter wave encounters an aircraft it is re-radiated or *reflected* from the aircraft and may cause *another* input to the receiver aerial (Fig 3). The total input to the receiver is the combination of the direct and reflected signals and

as the aircraft moves from one position to another the amplitude of the combined input varies as the two signals come into, and go out of, phase. This makes the picture flutter.

The fact that aircraft reflect e.m. waves to give such a noticeable effect on the screen of a cathode ray tube is important because enemy aircraft will also reflect waves which can then be used to detect their approach and determine their position so that fighters or missiles can be guided to intercept them.

How a Target is Detected

Imagine that you are on board a ship which is sailing in dense fog through a region where there are many icebergs. How can an iceberg ahead of the ship be detected?

One well-known method is to sound a short blast on a fog-horn, which directs the sound ahead of the ship, and listen for *echoes* (Fig 4). As the ship moves forward the horn is sounded at regular intervals and the crew listen for echoes in the pauses between the pulses of sound. Reception of an echo indicates an object ahead of the ship.

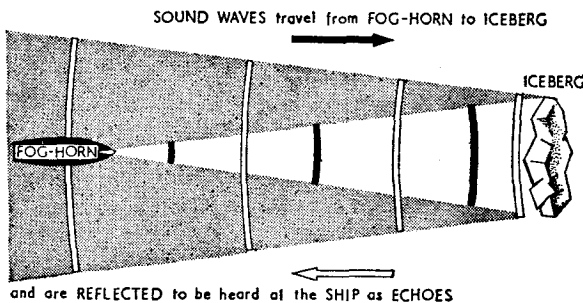


FIG 4. SOUND WAVES USED TO DETECT AN OBJECT

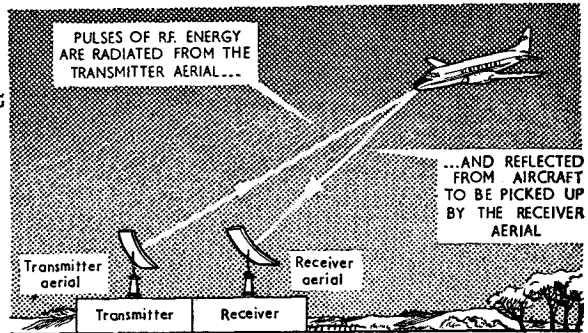


FIG 5. RADAR DETECTION OF AN OBJECT

Pulse-modulated radar detects aircraft in a similar manner using radio waves instead of sound waves (Fig 5). The transmitter is switched on for a *very short period only* to give a *pulse* of r.f. energy which is radiated from a directional aerial. If the e.m. wave encounters an aircraft, some of the energy is reflected back as an *echo* towards the transmitter-receiver. The output from the receiver is displayed on an *indicator* which shows the operator that there is an aircraft within range.

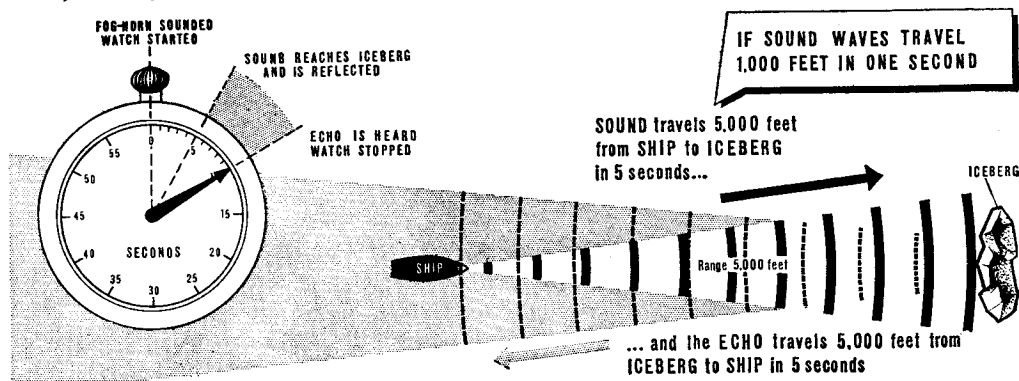
Several hundred pulses per second are normally transmitted and the receiver 'listens' for echoes in the intervals between pulses.

How Range is Measured

Let us think again of the ship sailing in fog. When the presence of an iceberg somewhere ahead of the ship is indicated by the reception of an echo the captain must find out how far ahead it is, i.e. its *range*, in order to take the necessary avoiding action.

As the speed of sound waves is known (approximately 1,000 feet per second) the range may be found by using a stop-watch to measure the time interval between the sounding of the fog-horn and the reception of an echo. Let us assume that the horn is sounded and an echo heard ten seconds later (Fig 6). In this time the sound travels 10,000 feet. This is the *total distance* from the ship to the iceberg and back to the ship. The *range* of the iceberg is thus:

$\frac{10,000}{2} = 5,000$ feet. This gives a simple relationship:



$$\text{Range} = \frac{\text{Speed of sound waves} \times \text{Time interval}}{2}$$

where range is in feet,
speed is in feet per second,
time interval is in seconds.

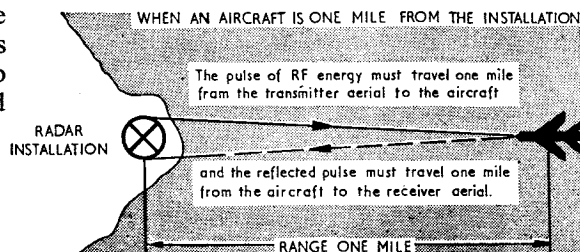
A pulse-modulated radar system uses radio waves in a manner similar to that above to measure the range of a target. Let us assume that an aircraft is exactly *one mile* from the radar installation, as in Fig 7, and that we wish to find the time interval between transmission and reception of the pulse.

The speed of e.m. waves is known to be 186,000 miles per second (or 300×10^6 metres per second). Therefore the time taken to travel from the transmitter to the aircraft and back is:

$$\frac{1}{186,000} + \frac{1}{186,000} = \frac{1}{93,000} \text{ second.}$$

Expressed in *microseconds* (μs) this is:

$$\frac{1}{93,000} \times 1,000,000 = \frac{1,000}{93} = 10.75 \mu\text{s}.$$



This figure is known as the *time of one radar mile*. For each mile of range the time interval is 10.75 microseconds. Thus if an aircraft is at 20 miles range the time interval is $20 \times 10.75 = 215$ microseconds.

In practice, of course, we work the other way round. The time interval is measured and from this we find the range. This can be found from the relationship mentioned earlier:

$$\text{Range} = \frac{\text{Speed of e.m. waves} \times \text{Time interval}}{2}$$

The units used depend upon whether the range is required in miles or in metres. If the range is required in miles, speed is given in miles per second (186,000) and time in seconds; if the range is in metres, speed is in metres per second (300×10^6) and time in seconds.

Examples *a.* Time interval between transmission of pulse and reception of echo is 320 microseconds.

$$\text{Range (in miles)} = \frac{186,000 \times 320 \times 10^{-6}}{2} = 29.76 \text{ miles.}$$

b. Time interval of 8 microseconds.

$$\text{Range (in metres)} = \frac{300 \times 10^6 \times 8 \times 10^{-6}}{2} = 1,200 \text{ metres.}$$

In certain ground radar stations the equipment is calibrated in *data miles* (6,000 ft.) instead of in statute miles (5,280 ft.). Similarly, most airborne radar navigational equipments are calibrated in *nautical miles* (6,080 ft.) and *knots* instead of in statute miles and m.p.h. From the known equivalent time for one radar mile ($10.75\mu\text{s}$), the times for one radar data mile ($12.2\mu\text{s}$) and one nautical mile ($12.36\mu\text{s}$) may be calculated.

How Range is Indicated

We have seen how a stop-watch measures the time interval between the blast from a fog-horn and the reception of the echo from an object ahead of the ship. The process may be simplified by using a stop-watch which is marked not in seconds but directly in feet of range. The exact

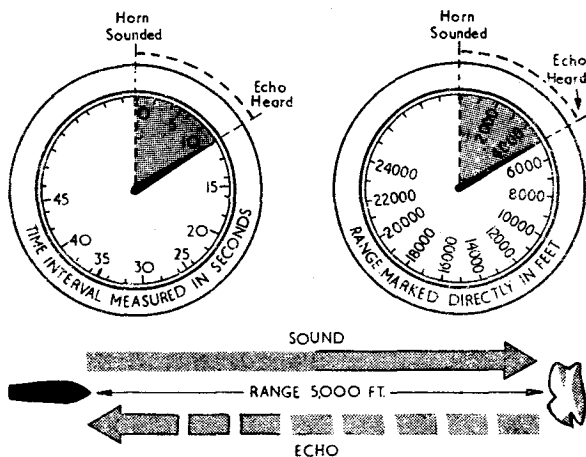


FIG. 8. INDICATION OF RANGE USING A STOP-WATCH

time interval corresponding to 1,000 feet range is calculated and the face of the watch is marked at that point as 1,000 feet. Similar marks are made to indicate ranges of 2,000 feet, 3,000 feet and so on (Fig. 8).

In radar measurement of range the time intervals to be measured are *very short* (a few microseconds). Such short time intervals cannot be measured by a stop-watch and so we use a *cathode ray tube* (c.r.t.) to produce a spot of light which can be moved over the screen to build up a pattern or a picture representing the inputs to the deflecting system. Let us now see how these principles are applied to radar.

The horizontal trace on an electrostatic c.r.t. is obtained by applying a saw-tooth timebase waveform to the X-deflecting plates of the c.r.t. In radar, the timebase waveform must be 'synchronized' with the transmitter operation so that the spot starts to move across the screen at the instant the pulse of r.f. energy leaves the aerial (Fig. 9). The spot moves at an *exactly known speed*, e.g. it may move two inches in the time that the transmitted pulse moves twenty miles through space (see Fig. 9).

Suppose that after travelling 20 miles the pulse is reflected by an aircraft. The echo travels back 20 miles to the radar installation so that the pulse and the echo make a combined journey of 40 miles. In this time the spot will have moved four inches. If the received echo is converted instantly by the receiver into a voltage pulse which is applied to the Y-deflecting plates of the c.r.t. the spot is deflected upwards for the duration of the pulse, making a deflection 'blip' four inches from the starting point of the trace. This 'blip' indicates a target at a *range* of 20 miles.

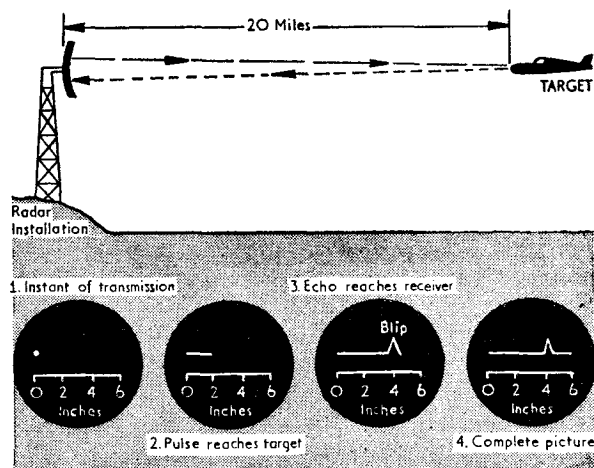


FIG. 9. FORMATION OF RANGE DISPLAY

If the spot always moves at the same rate, every inch of trace represents five miles of range. The range of any target which produces an echo can therefore be found simply by measuring the distance of its 'blip' from the start of the trace. One fleeting 'blip' would be of little value as a means of indicating a target. We therefore arrange for a continuous picture to be built up by sending out pulses in rapid succession.

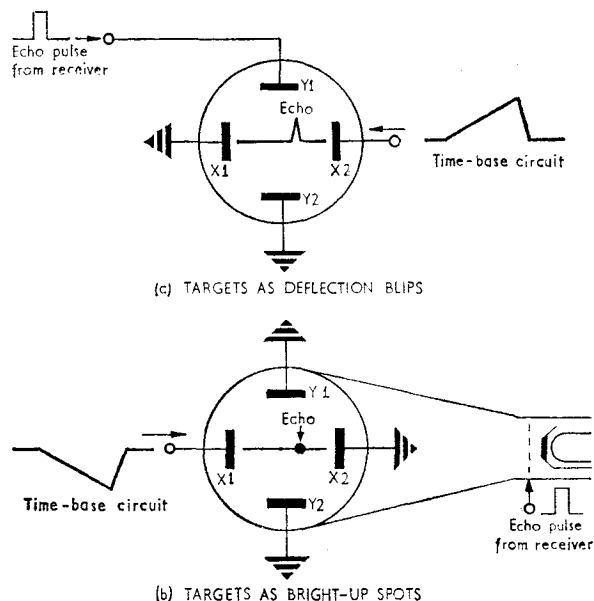


FIG. 10. METHODS OF TARGET INDICATION

The echo blips are produced by converting the signals picked up by the receiver into short pulses of d.c. voltage which we apply to the Y-deflecting plates of the c.r.t. This pulls the spot upwards for the duration of each pulse (Fig. 10a). Sometimes the echo is required to show up as a

bright spot on the trace instead of as a blip. In this case the receiver output is applied as short positive-going pulses of voltage to the *grid* of the c.r.t. We know that the brightness control of an oscilloscope varies the voltage on the grid of a c.r.t. thus altering the brightness of the display. Therefore a target echo applied as a *positive* pulse to the *grid* of the c.r.t. increases the brightness of the spot on the trace for the duration of each pulse (Fig 10b). This 'bright-up' form of display is known as *intensity modulation* and it can also be obtained by applying *negative* pulses to the *cathode* of the c.r.t.

As soon as the spot has covered the full width of the c.r.t. screen it is returned to the left-hand side and remains there until the transmitter 'fires' again. The whole process is then repeated. A radar transmitter may produce many hundreds of pulses per second. If it 'fires' 500 times in one second (a typical figure) the spot moves over the trace 500 times each second. This recurring movement of the spot ensures that a bright trace is built up and maintained. The repetition rate is too fast for the eye to be able to follow the separate movements of the spot and we see the trace as an apparently steady line on the screen, with the target echoes moving slowly along the trace as the range of each target changes. This type of display is known as a *type A* display and it shows *range* only. Fig 11 illustrates the sequence of events in each cycle of operation.

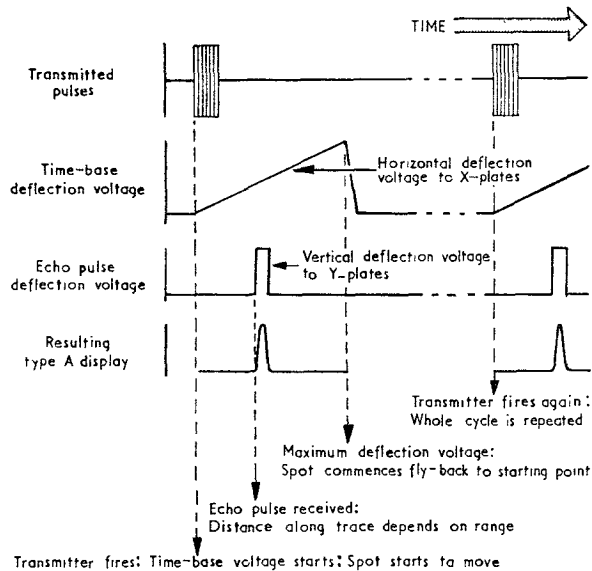


FIG 11. ONE CYCLE OF OPERATION IN PRODUCTION OF TYPE 'A' DISPLAY

Changing the Range Scale

For the display described in the preceding paragraphs the maximum range which can be shown on a six-inch c.r.t. is 30 miles. If, on the same c.r.t., we want to check for echoes from a range *greater* than 30 miles we *reduce* the speed of the spot, i.e. the *velocity* of the timebase is reduced (see later). For example, if the speed is *halved* the spot travels only *one inch* in the time taken for an echo to return from a target at ten miles range, and the maximum range which can then be shown on a six-inch c.r.t. is 60 miles.

Notice in Fig 11 that the time during which echoes may be displayed on the c.r.t. is limited to the *interval between transmitted pulses*. The number of pulses that can be transmitted per second, i.e. the *pulse recurrence frequency* or the *pulse repetition frequency* (p.r.f.), must therefore be such that the interval between pulses is sufficiently long to allow the spot to complete each trace and return to the starting point.

How Bearing is Obtained Using Sound Waves

Let us return to the ship sailing through fog and imagine that the captain has used the fog-horn to *detect* an iceberg somewhere ahead of the ship and has measured its *range*. Earlier we assumed that the fog-horn would beam the sound in a generally forward direction ahead of the ship. An iceberg situated anywhere within the sound beam will cause an echo but, as shown in Fig 12, the captain cannot be sure of its *exact* position.

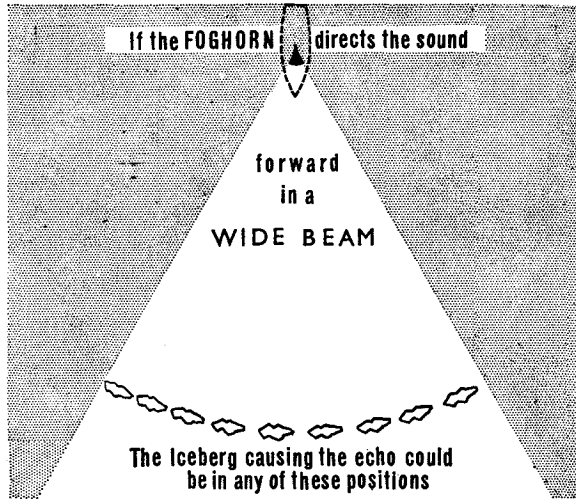


FIG 12. NEED TO DETERMINE BEARING

To take the necessary avoiding action the captain also needs to know the *relative bearing* of the iceberg, i.e. whether it is dead ahead or to one side of the direction in which the ship is heading.

In order to state the bearing clearly we imagine that the sea ahead of the ship is marked out as in Fig 13. The captain can obtain a more accurate bearing by using a fog-horn which concentrates the sound waves into a *very narrow beam*. Echoes are then heard only when the beam is aimed directly at the iceberg and the direction in which the horn is pointing gives the *relative bearing*.

Notice that while the narrow beam of sound is aimed at iceberg A there is no echo from iceberg B which is much nearer to the heading of the ship. To ensure that all icebergs ahead are detected the horn must first be pointed well to one side and sounded while the crew listen for echoes, then turned slightly towards the heading and sounded again and the process repeated until the horn is pointing well to the other side. This procedure is known as *scanning* (Fig 14) and it should be a continuous process.

Radar Bearings

The bearing of a target relative to a radar installation is obtained in a similar manner. To find the bearing the pulses of r.f. energy produced by the transmitter are concentrated into a very narrow beam by means of a *directional aerial array*.

To *scan* the whole of the area from which the aircraft might approach, the aerial assembly is systematically turned from side to side (Fig 15). Echoes are received only when the beam is aimed directly at the target and, since we know the direction in which the aerial is pointing at any given time, we can find the bearing of any aircraft which reflects the r.f. pulses.

For many applications the radar aerial is required to rotate through a full 360° so that the complete area surrounding the radar installation may be examined. In such cases it is usual to

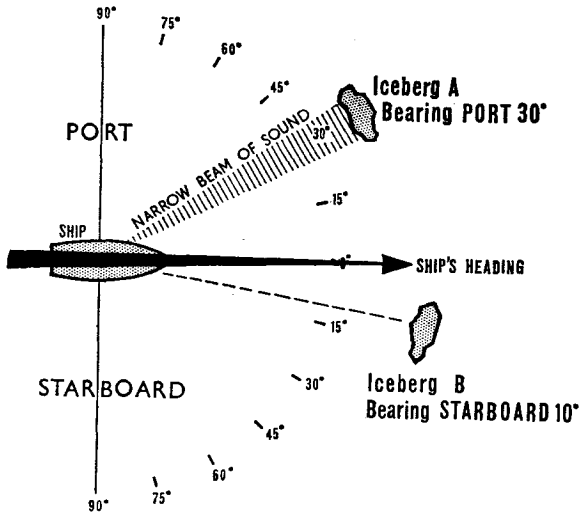
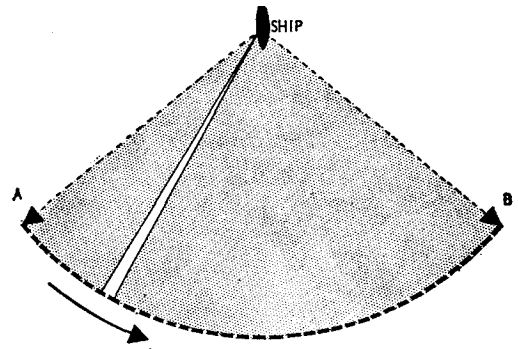


FIG 13. DETERMINATION OF BEARING



TO 'SCAN' A WIDE ANGLE AHEAD OF A SHIP WITH A NARROW BEAM OF SOUND, THE BEAM IS FIRST POINTED IN DIRECTION 'A', THEN TURNED SLOWLY THROUGH THE WIDE ANGLE TO BE SCANNED UNTIL IT IS POINTING IN DIRECTION 'B'.

FIG 14. BASIC IDEA OF SCANNING

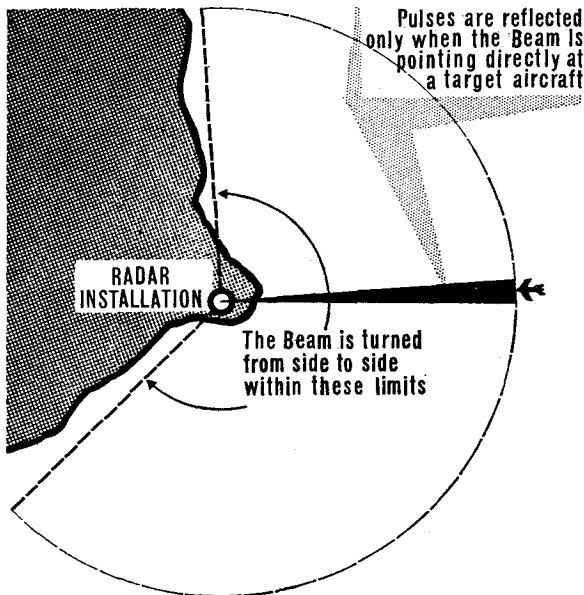


FIG 15. RADAR BEARINGS

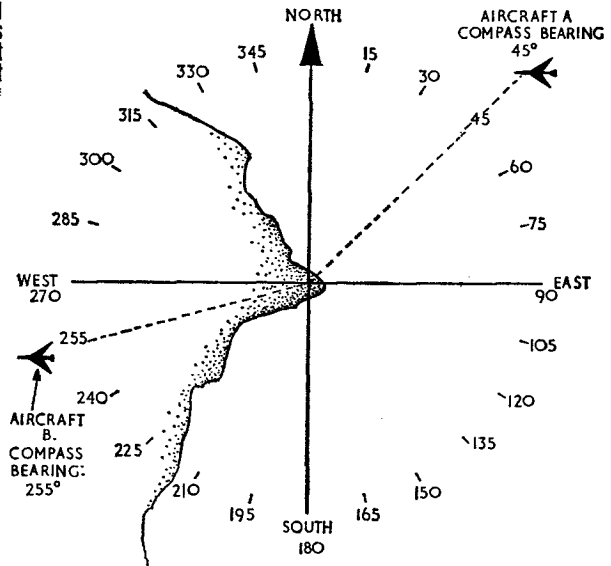


FIG 16. COMPASS BEARING OF TARGET

measure the *compass bearing* of the target. North is used as the reference direction and bearings are given in degrees relative to North (Fig 16).

When the radar equipment is carried in a fighter aircraft the bearing of a target is given in degrees to port or starboard of the direction in which the fighter is heading (Fig 17).

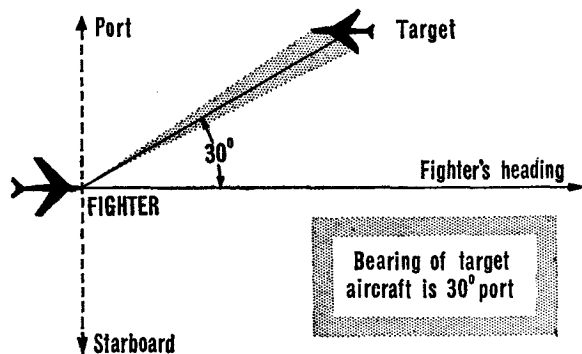


FIG 17. RADAR BEARING FROM AN AIRCRAFT

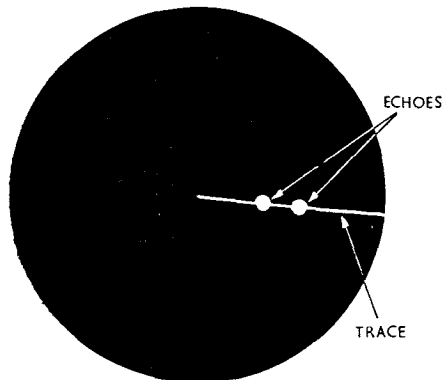


FIG 18. TRACE AS THE 'HAND OF A CLOCK'

Indication of Bearing

A range trace, such as that produced by a type A display, if made to start at the *centre* of the c.r.t. and extend to the edge, will look like the hand of a clock. If at the same time echoes are made to appear as 'bright-up' spots the result will be as shown in Fig 18.

If the trace could be made to point towards 'twelve o'clock' when the radar aerial is pointing *due North* then any echoes appearing on the trace must be from targets *due North*.

When the aerial is turned towards the East any echoes appearing on the trace must then be from targets East of North. If at the same time the trace could be moved *clockwise* by the *same amount* as the aerial moved *towards the East* the new echoes would appear in their *correct relative position* on the screen. By continuing this process right through 360°, the trace moving *in synchronism* with the aerial, the trace on the c.r.t. will look like the hand of a clock rotating steadily with the aerial and the echoes will appear as bright spots at appropriate parts of the

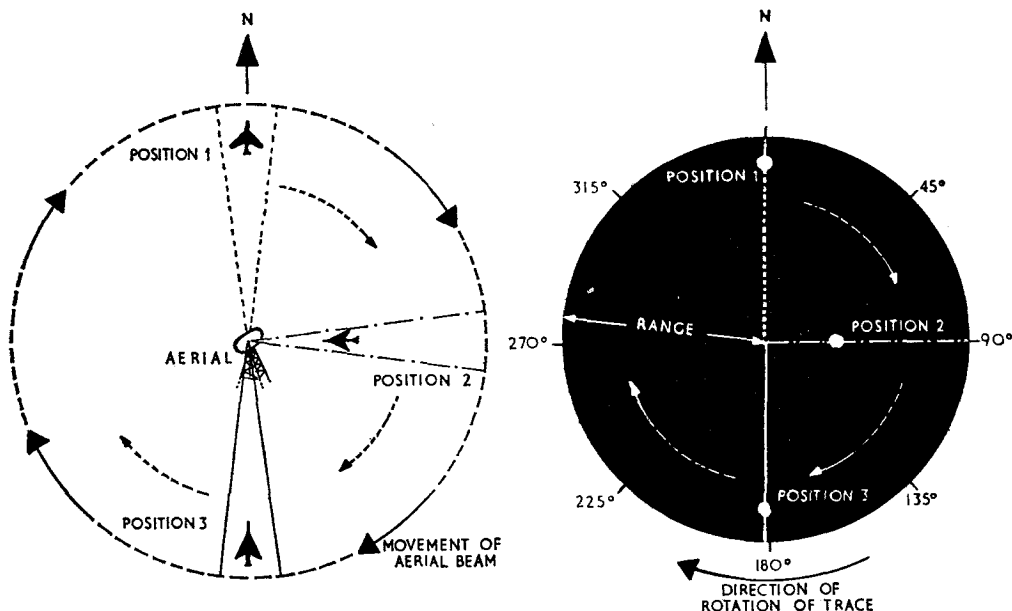


FIG 19. THE PLAN POSITION INDICATOR

screen. This arrangement is illustrated in Fig 19. To ensure a bright, clear picture the c.r.t. screen has a long 'afterglow'. Thus when an echo is 'painted' it remains bright until the trace comes round again to re-paint it.

The display described is known as a *plan position indicator* (p.p.i.). The *distances* of the echoes from the *centre* of the screen indicate the *ranges* of targets, and the *angular position* of each echo indicates the *bearing* in plan or '*in azimuth*' of the corresponding target. Such a display is in effect a 'radar map'. Fig 20 shows a typical picture on the p.p.i. of a ground search radar.

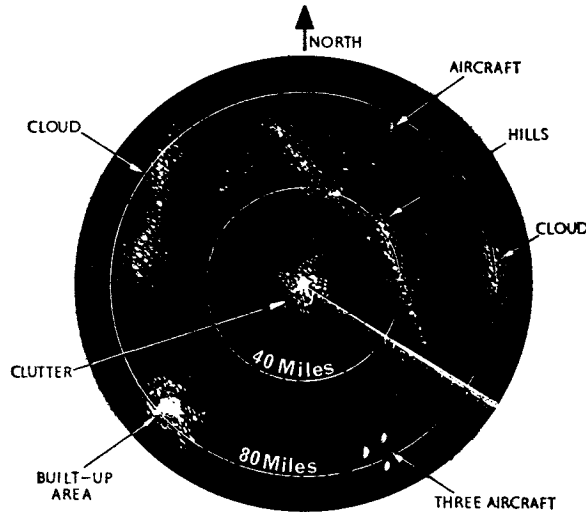


FIG 20. TYPICAL PPI DISPLAY

How the Height of a Target is Determined

So far we have seen how a target can be *detected* and how its *range* and *azimuth bearing* can be found. We are thus a long way towards pin-pointing the position of a target in space. The only factor missing, if the target is an aircraft or a missile, is the *height* of the target. A radar installation which indicates range, bearing and height of a target gives, in effect, a three-dimen-

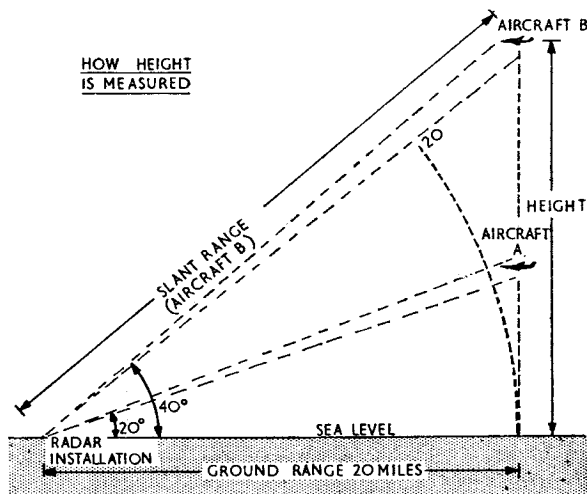


FIG 21. SLANT RANGE AND ANGLE OF ELEVATION

sional picture. Such radars are in fact often called 3-D radars. It is not usual however to show all the factors on a single c.r.t. display. Normally two c.r.t.s are used, one to show range and bearing (e.g. a p.p.i.) and the other to show range and height.

To find the height of a target we use an aerial array which can swing the radiated beam up and down. Fig 21 shows two aircraft which are both at a *ground range* of 20 miles but are flying at different heights. Two important facts can be obtained from the diagram:

a. The *actual distance* between the radar and aircraft B is much greater than that between the radar and aircraft A. The distance to the target in each case is called the *slant range*.

b. With the radar beam directed at aircraft A the angle between the beam and the horizon (the *angle of elevation*) is 20° . For aircraft B it is 40° . Thus the greater the height of an aircraft, at a given ground range, the greater is the angle of elevation.

We shall now see how these facts are used to find the height of an aircraft and its true ground range. The known factors are slant range r and angle of elevation θ of the aerial. From Fig 22:

$$\sin \theta = \frac{\text{Height}}{r} \therefore \text{Height} = r \sin \theta.$$

$$\cos \theta = \frac{\text{Ground range}}{r} \therefore \text{Ground range} = r \cos \theta.$$

Example. If the slant range of an aircraft is measured as 30 miles and the angle of elevation as 5° :

$$\text{Height} = 30 \sin 5^\circ = 30 \times 0.0872 = 2.6 \text{ miles or } 13,720 \text{ feet.}$$

$$\text{Ground range} = 30 \cos 5^\circ = 30 \times 0.9962 = 29.89 \text{ miles.}$$

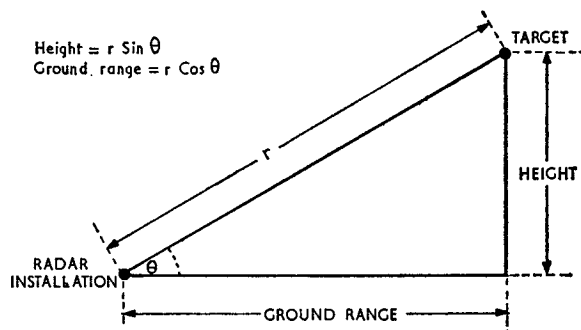


FIG 22. CALCULATION OF HEIGHT AND GROUND RANGE

How Height is Indicated

In practice it is not necessary to calculate the height and ground range of each target because modern radar equipment includes a suitably designed indicator which shows these factors directly.

The aerial used for height-finding radiates a beam which is narrow in the vertical direction but which spreads horizontally (Fig 23a). It is first turned towards the azimuth bearing of the target (obtained from the p.p.i.) and then it *nods up and down* between the limits to be scanned.

The indicator used with this aerial has a form of bright-up (intensity-modulated) display. The trace is formed in much the same way as for a p.p.i., the trace moving in synchronism with the aerial; but since the aerial swings in a *vertical arc* the illuminated part of the c.r.t. is a sector in elevation (Fig 23c). Ground range and height are pre-calculated and are marked on a scale placed over the face of the c.r.t.

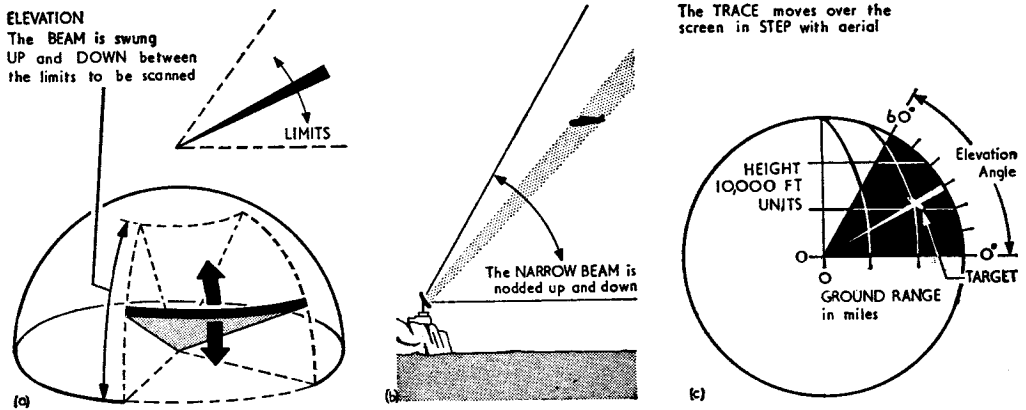


FIG 23. INDICATION OF HEIGHT AND GROUND RANGE

Conclusion

In this chapter we have seen that pulses of r.f. energy radiated by a radar transmitter may be reflected from aircraft. The resulting echoes can be used to *detect* and to find the *range*, *bearing* and *height* of enemy aircraft so that fighters or missiles can be guided to an interception position in the sky.

We must now find out how these pulses are produced and radiated and how the reflected pulses are received and displayed in the radar installation.

CHAPTER 2

BASIC REQUIREMENTS OF A PULSE-MODULATED RADAR SYSTEM

Introduction

We have seen that a primary radar installation must be able to detect targets at a distance and also provide information on their range, bearing and height. To be able to do this satisfactorily a pulse-modulated radar must have the units shown in Fig 1.

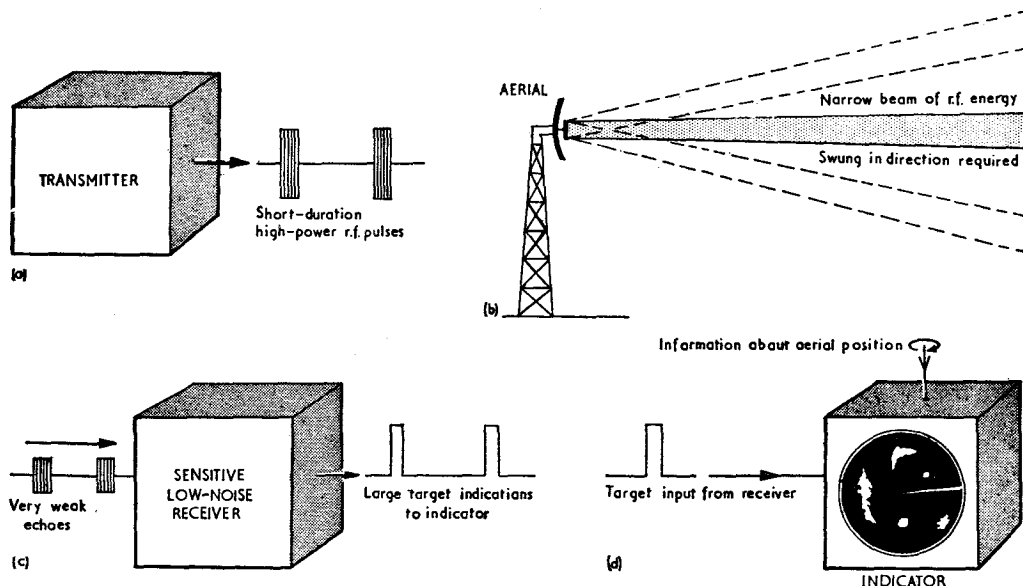


FIG 1. BASIC UNITS OF A PULSE-MODULATED PRIMARY RADAR

- A *transmitter* capable of producing short-duration, high-power pulses of r.f. energy at a given repetition frequency.
- An *aerial system* producing a very narrow beam of r.f. energy which may be used for scanning.
- A *sensitive receiver* capable of receiving and amplifying the very weak echoes for application to the indicator.
- An *indicator* which measures time intervals of only microseconds duration for range measurement, and which is connected to the aerial system to give indication of the bearing or the height of the target.

Let us now examine some of the stages in a primary radar installation. We can then go on to link the various units together to form a block schematic diagram of a basic system.

The Pulse-modulated Transmitter

The transmitter pulses are produced at a regularly recurring rate, the *number* of pulses produced each second being known as the *pulse recurrence (or repetition) frequency* (p.r.f.), measured in pulses per second. The *length of time* for which the transmitter is switched on to give each pulse is known as the *pulse duration*, measured in microseconds. Alternative names for pulse duration are *pulse width* and *pulse length*.

A radar transmitter also operates at a very high frequency and, for the duration of each pulse, produces a very high peak power output. Frequencies up to 30,000 Mc/s are common and peak powers of 1MW and more are used. The high powers are necessary to ensure adequate 'illumination' of the target so that a good echo may be received. We shall see the reason for the use of the high frequencies later.

The system used to produce the high-frequency, high-power pulses of short duration at a required repetition rate is illustrated in block schematic form in Fig 2.

a. Master timing unit. This unit produces timing pulses which recur at precise and regular intervals of time and so *determines the p.r.f.* of the equipment. The indicator time-base is synchronized with the transmitter pulses by applying 'sync pulses' from the master timing unit to the indicator timebase. This causes the trace to move across the screen of the c.r.t. at the instant the transmitter fires each pulse.

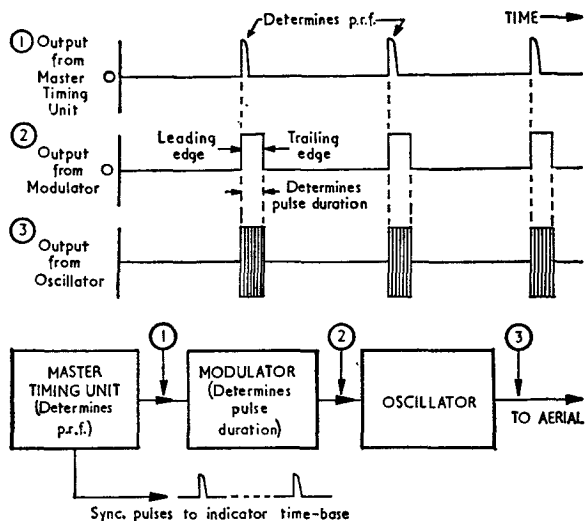


FIG 2. THE PULSE-MODULATED TRANSMITTER

b. Modulator. Because of the high rate of switching (many hundreds of pulses per second) and the very short time intervals being used (a few microseconds at the most for the pulse duration) the transmitter operation cannot be controlled by normal switches or relays. The circuit which does this switching, and also supplies the input power required by the oscillator, is the *modulator*. It is an electronic circuit which is 'triggered' by the output from the master timing unit and which produces a *d.c. pulse* whose *duration* is determined by the circuitry of the modulator. This d.c. pulse of *controlled pulse duration*, recurring at the precise instants of time determined by the master timing unit, is used to switch the oscillator on and off.

c. Oscillator. This unit generates the high-frequency oscillations at the high power required. It is switched on by the rising or leading edge of the d.c. pulse from the modulator and is switched off by the falling or trailing edge of the pulse. Thus the transmitter produces a *pulse of r.f. energy* at the frequency of the oscillator. We shall see later that to produce the very high frequencies at the high powers needed by centimetric radars the oscillator uses special microwave devices (e.g. magnetrons or klystrons). Note also that although the pulse duration may be very short the frequency is sufficiently high to ensure that each pulse con-

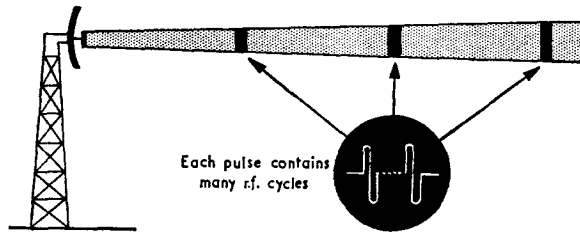


FIG. 3. RADIATED PULSES

tains a *large number of cycles* of radio frequency (Fig 3). For a frequency of 3,000 Mc/s and a pulse duration of $1\mu s$ each pulse contains 3,000 cycles.

Many factors have to be considered before the p.r.f. and the pulse duration of a particular radar installation are decided, but typical figures are a p.r.f. of 500 pulses per second and a pulse

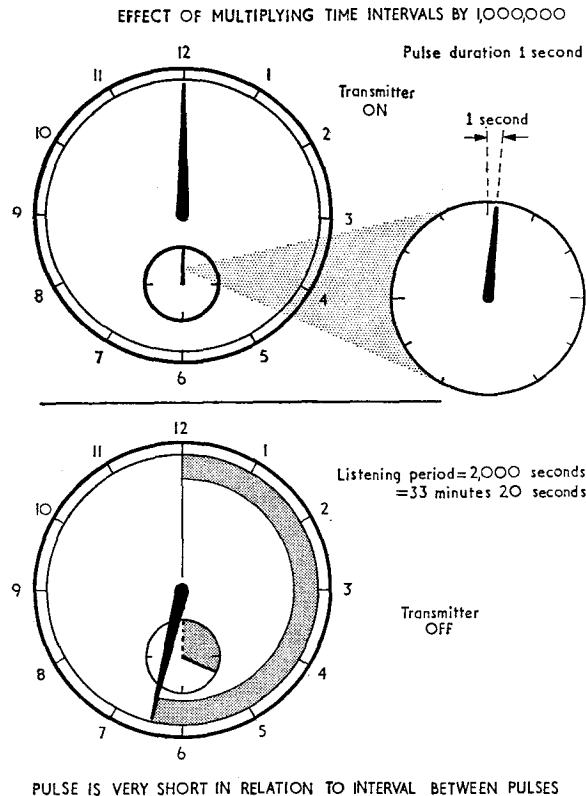


FIG. 4. TIME INTERVALS IN RADAR

duration of $1\mu s$. For the figures quoted the transmitter 'fires' for only *one millionth* of a second and is then off for $\frac{1}{500}$ second or 2,000 *millionths* of a second.

To give us an idea of what these figures really mean let us for the moment imagine them to be multiplied by one million as in Fig 4. This emphasises that the transmitter fires for only a *very short* period of time, the interval between each pulse being *relatively long*.

Aerial System

Aerial. To locate an object in space we need an aerial capable of producing a *narrow beam* of energy so that the high peak power output from the transmitter during each pulse may be concentrated to cover a small region of space and so that accurate bearings in both azimuth and elevation may be obtained. To produce such a beam some form of *aerial array* is required and, as we have seen in Part 1 of these notes, the *higher* the frequency the *smaller* are the aerial elements in the array. At v.h.f. and above, the physical size of the aerial elements are such that we are able to construct small aerial arrays capable of rapid movement.

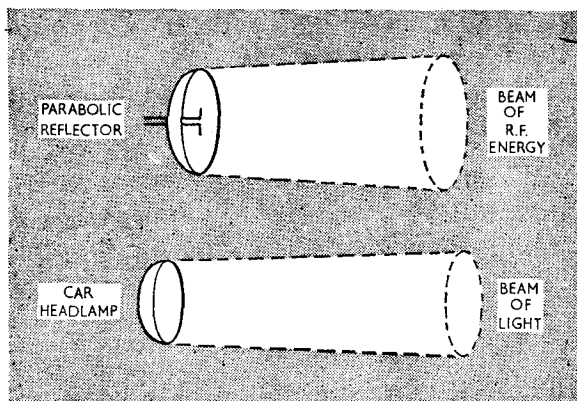


FIG 5. CONCENTRATION OF ENERGY IN A BEAM

Centimetric radar uses frequencies in the band 3,000 to 30,000 Mc/s. At these super-high frequencies the aerial assemblies used for radiating a narrow beam are very similar to the structures used for focusing light rays. Typical of these is the *parabolic reflector* which may be compared with the headlight of a car (Fig 5). We shall learn more about the construction and operation of such aerals in later chapters of this book. It is sufficient to know at this stage that centimetric radar aerals are small, that they produce a narrow beam of r.f. energy and that they can be rotated and tilted fairly easily.

Scanning. Locating an object in space with a narrow beam of r.f. energy is rather like looking for a needle in a haystack. The beam must be capable of being swung in any required direction and the search for an object in a given volume of space must be carried out systematically to ensure that the whole volume is covered. To do this the aerial beam is made to *scan* the whole region which is to be investigated. Fig 6 shows one very simple form of scanning. A small overlap of the beam between adjacent horizontal scans is usual to ensure complete coverage. When the aerial is pointing maximum right at the minimum angle of elevation to be used the movement is reversed and the aerial returns to its original position in a similar manner. The scanning is repeated continuously to give the operator a complete indication of all targets within the region which is being scanned. Other methods of scanning are used and these are considered in more detail later.

T-R switch. Most pulse-modulated primary radars use an aerial that is *common* to both transmitter and receiver. This may be done because while the transmitter is working we do not need to use the receiver, and while the receiver is working the transmitter is switched off. We can therefore connect the *transmitter* to the aerial for the *duration* of each pulse and connect the *receiver* to the aerial for the *interval between pulses*.

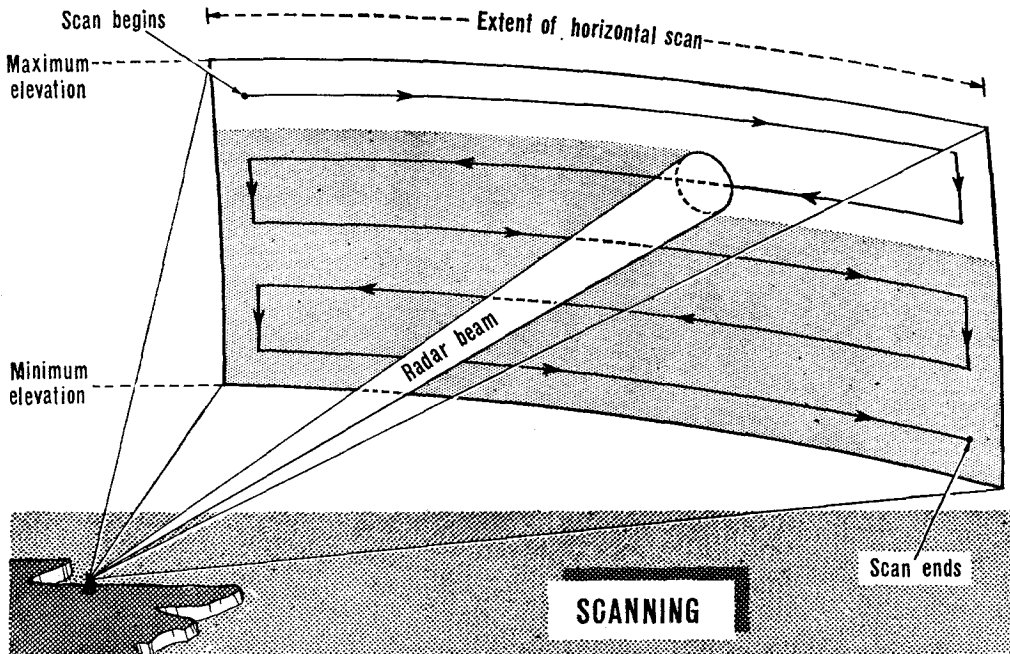


FIG 6. BASIC IDEA OF SCANNING

This may be done by the transmit-receive (T-R) switch, the action of which is illustrated in Fig 7. Although the T-R switch is represented here by the conventional symbol for a switch, in practice it is an electronic device which operates automatically. Normal mechanical switches or relays cannot be used because of the very high rate of switching.

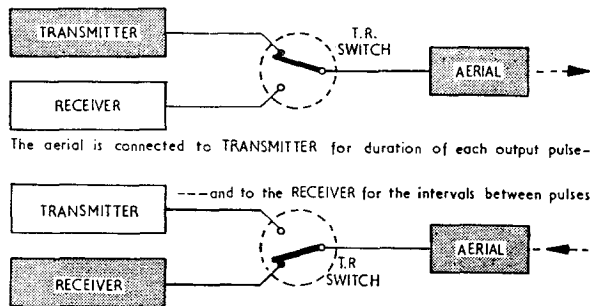


FIG. 7. T-R SWITCHING

Output to indicator. We have seen that information on the angular position of the aerial is conveyed to a p.p.i. for indication of bearing in azimuth, and information on the angle of tilt is conveyed to a range-height indicator for indication of bearing in elevation. This is done by one of the remote position indication systems discussed in Part 1 of these notes, i.e. Desynn, M-type or synchro systems. A typical arrangement is illustrated in Fig 8.

Receiver

The receiver amplifies the reflected pulses picked up by the aerial. After amplification the pulses are demodulated and the resulting d.c. pulses are applied to the indicator c.r.t. for

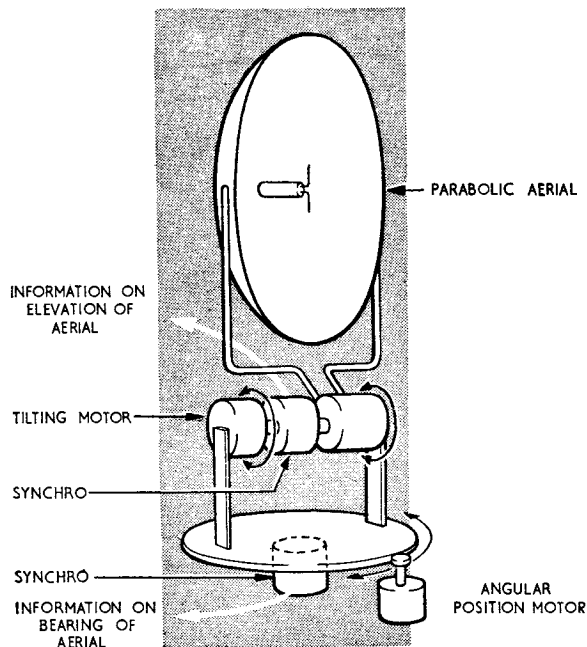


FIG 8. TRANSMISSION OF AERIAL POSITION INFORMATION TO INDICATORS

display of target information. The received echoes are very weak and may be only a few microvolts. The input voltage required by the indicator however to give a reasonable display may be 10 volts or more. Thus the receiver must have a *high gain*.

As we shall see later there is a limit to the amount of useful gain which can be obtained in practice. The main limitation is *noise*. The interference we get in television in the form of random flashes of light on the screen is the result of *unwanted* electrical variations which have either entered the receiver circuit from an outside source (e.g. the unsuppressed ignition system of a motor car) or been generated within the receiver itself. This interference is referred to as noise. The main sources of noise within a receiver are mentioned in Part 1 of these notes and will be dealt with in more detail later.

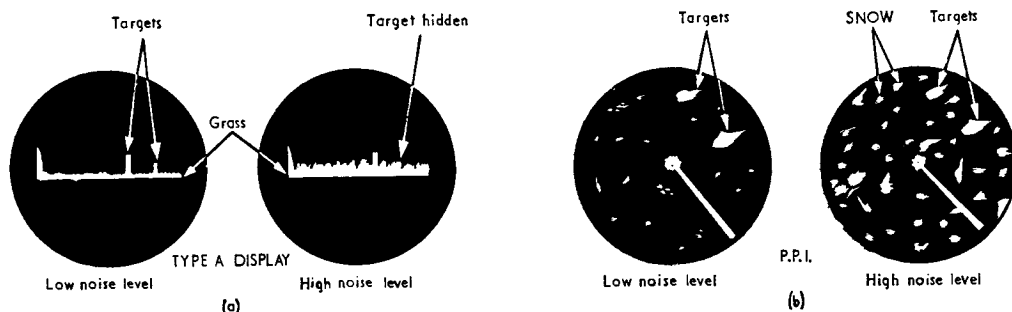


FIG 9. EFFECT OF NOISE ON INDICATOR DISPLAYS

It is sufficient for us to know at this stage that the noise level of a receiver has a marked effect on the performance of the equipment. If the indicator uses a simple type A display (Fig 9a) the noise voltages are applied to the Y-deflecting plates of the c.r.t. together with the signal echoes to give the effect which we call 'grass' on the trace. If the noise level is sufficiently high

the 'grass' may *hide* the weak echoes from targets at long range. In a bright-up (intensity-modulated) type of display, such as a p.p.i., the signal and noise voltages are applied together to the c.r.t. grid (or cathode). Noise then appears as random and confusing flashes of light on the c.r.t. screen (Fig 9b), an effect sometimes called 'snow'.

We must therefore take precautions to keep the noise generated by the receiver to an absolute minimum. It is pointless having a high-gain receiver if it also has a high noise level; the noise is merely amplified further.

One further point may be noted about the receiver at this stage: to preserve the *shape* of the reflected pulse the receiver circuits must have a *wide bandwidth*. This follows from the fact that a square or rectangular pulse of voltage contains a very wide band of frequencies which must be accepted by the receiver if distortion is to be avoided. A distorted pulse on the c.r.t. display makes it difficult to determine range accurately.

The Indicator Unit

The indicator is used to show the operator the range, bearing and height of all aircraft within range of the radar. The relevant information is then passed on to the controller. Very often only the *range* of a target is required; a simple *type A* display is then sufficient. Where *range and bearing* are required a p.p.i. display is usually used. Where *range and height* are required a range-height indicator is used. Where all three co-ordinates are required it is usual to have two c.r.t.s—one a p.p.i. to show range and bearing and the other a range-height indicator.

We have seen in Chapter 1 how these displays are produced, and earlier in this chapter we showed how information concerning the attitude of the aerial is transmitted by synchro systems to the indicator. Fig 10 summarizes the main points.

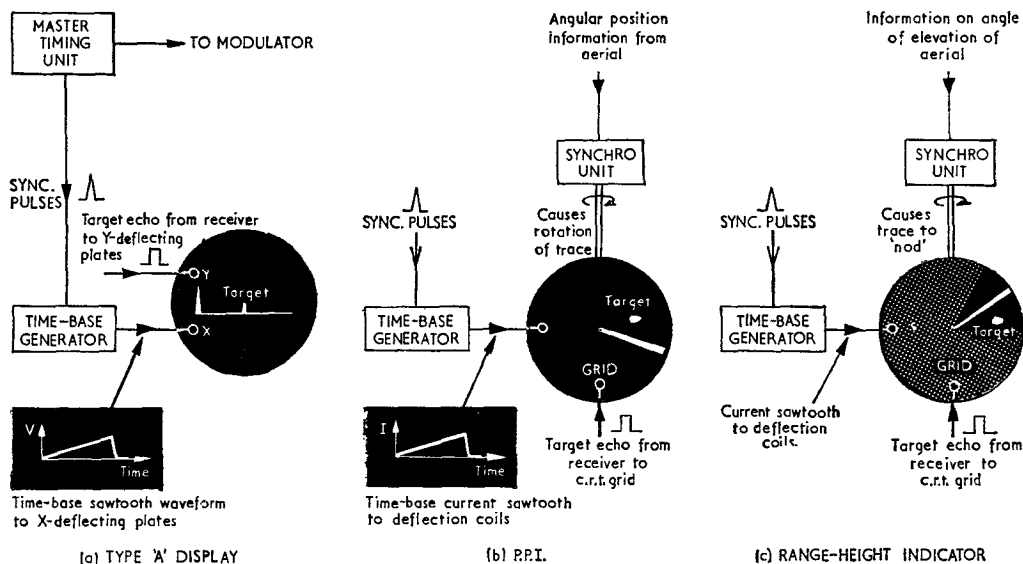


FIG 10. TYPE 'A', PPI AND RANGE-HEIGHT INDICATOR DISPLAYS

Other Radar Displays

A display can be designed to present almost any required information about a target in terms of range, bearing, height and elevation, or any two combinations of these. Fig 11 illustrates a few examples of additional displays in common use.

a. Type B. This shows bearing against range. The timebase voltage is applied to the *Y-deflecting plates* so that the range trace is *vertical*. The position of the trace in the horizon-

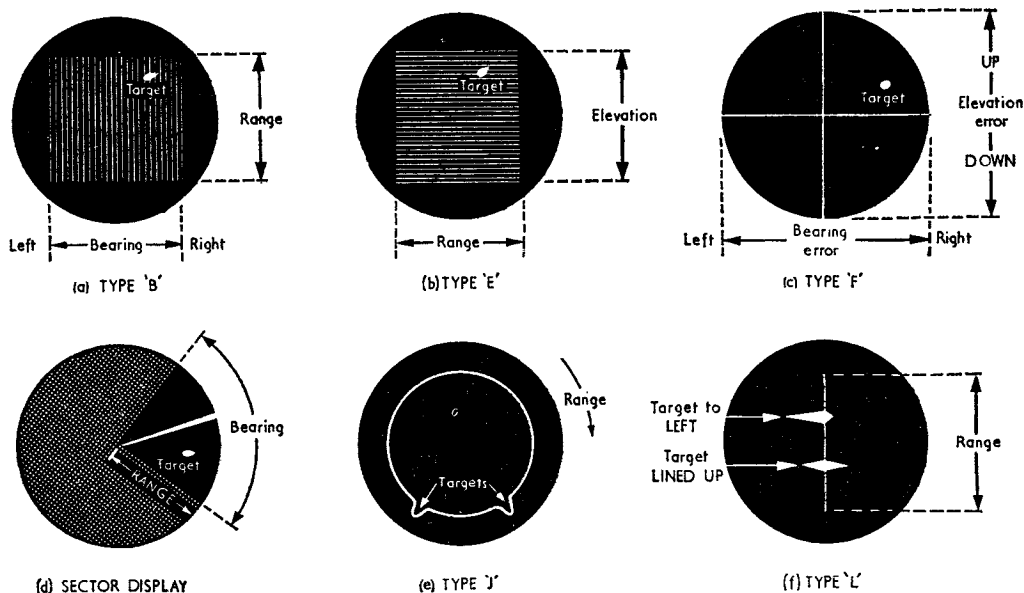


FIG 11. OTHER COMMON RADAR DISPLAYS

tal direction corresponds to the direction in which the aerial is pointing at that instant and each vertical trace moves across the screen in step with the aerial. Intensity modulation is used. The aerial scan is confined to a chosen 'sector', the limit of the swing often having a maximum value of $\pm 80^\circ$.

b. Type E. This shows range against elevation. The timebase voltage is applied to the X-deflecting plates and the position of the trace in the vertical direction corresponds to the angle of elevation of the aerial. Intensity modulation is used.

c. Type F. This shows bearing and elevation *errors* for a single target. A timebase voltage is not required because range is not being measured. DC error voltages are applied to the X-plates for bearing error and to the Y-plates for elevation error. When there is no error, i.e. aerial pointing directly at the object, the target appears as a bright spot at the intersection of the two crossed lines.

d. Sector display. This shows ranges and bearings of targets *over a given sector*. It is similar to the p.p.i. but since the aerial turns from side to side through an arc instead of turning a full circle the display is a *segment* of a full p.p.i. display. The sector display shows the same information as a type B but the type B shows more detail because it uses more of the c.r.t. screen area.

e. Type J. This is a modification of the type A display. It uses a circular timebase to indicate range only, the trace being π times longer than that for a corresponding type A display. A larger scale is thus obtained and targets close together in range can be more easily displayed. Deflection modulation is used.

f. Type L. This is a modification of the type A display. Two radar beams are used and the echoes due to each are placed back-to-back on the display. When the two blips are of equal size the aerial is pointing directly at the target. Bearing against range can be displayed in this way.

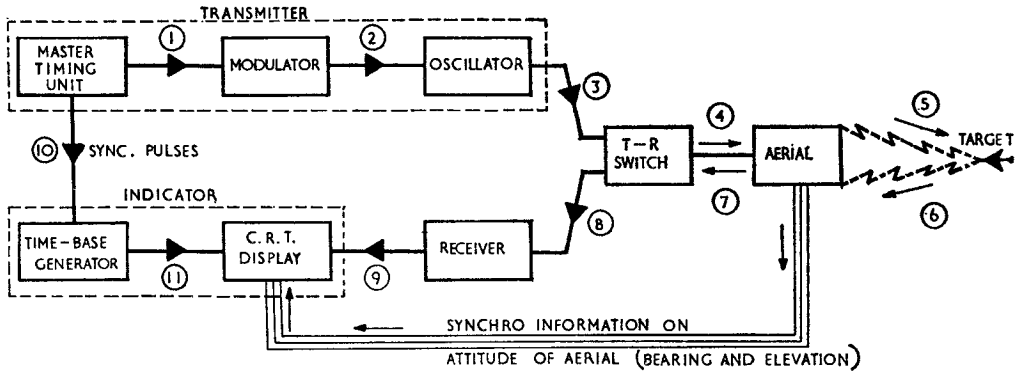


FIG 12. BASIC PRIMARY RADAR BLOCK SCHEMATIC DIAGRAM

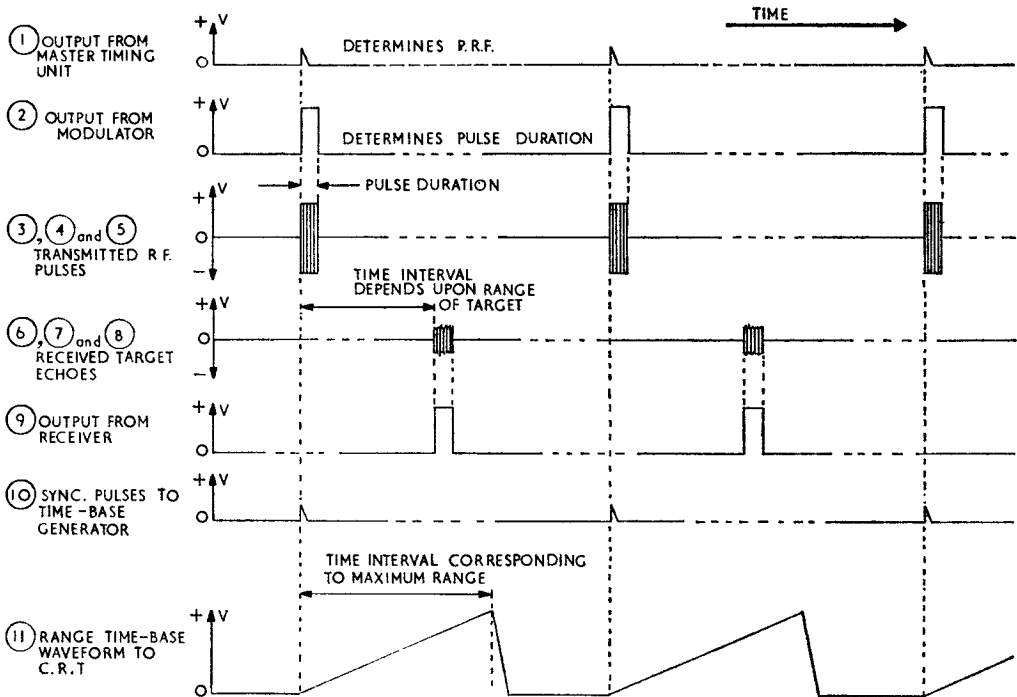


FIG 13. WAVEFORMS AT VARIOUS POINTS IN A PRIMARY RADAR SYSTEM

Schematic Diagram of Basic Primary Radar

Fig 12 shows how the various stages which we have discussed are linked together to form a block diagram of a basic primary radar system.

The waveforms which we should see if we connected an oscilloscope to the numbered points in Fig 12 are illustrated in Fig 13.

The action of each 'block' may be summarized as follows:

a. Master timing unit. The timing pulses produced by this unit *control the p.r.f.* of the equipment and are applied to:

- (1) The *modulator* to 'trigger' the transmitter operation at precise and regularly recurring instants of time.
- (2) The *indicator timebase generator* to synchronize the c.r.t. trace with the transmitter operation.

b. Modulator. This produces rectangular d.c. pulses of *known pulse duration* which switch the oscillator on and off.

c. Oscillator. This produces the very high frequency, high power output in pulses of short duration. The p.r.f. is determined by the master timing unit and the pulse duration by the modulator.

d. T-R switch. This automatically connects the transmitter to the aerial for the duration of each output pulse, and connects the receiver to the aerial for the intervals between pulses.

e. Aerial. This radiates the transmitter output in a *narrow beam* and picks up the reflected echoes for application to the receiver. The aerial may be moved for *scanning*, the movements being conveyed by synchros or servomechanisms to the indicator.

f. Receiver. This amplifies the very weak echoes and presents them in a suitable form for display on the indicator c.r.t.

g. Indicator timebase generator. This produces the range trace on the c.r.t. screen. The sync pulses from the master timing unit ensure synchronization of indicator and transmitter operations.

h. Indicator display. This presents the required target information in a suitable form.

Secondary Radar Systems

So far we have considered pulse-modulated *primary* radar systems which rely on the *reflection* of the incident energy by the target. Instead of relying on reflections we could use the pulses received from the primary radar to 'trigger' a *transmitter* carried by the target. Such systems are known as *secondary radar systems* and can only be used when the target is *friendly*. When used in this manner the primary radar transmitter is known as the *interrogator* and the secondary installation in the target is called the *transponder* (transmitter-responder).

One of the earliest applications of secondary radar was the identification equipment carried by all friendly ships and aircraft. In this system the primary radar on the ground is the interrogator and the ship or aircraft carries the transponder. When the transponder receives a signal of the correct frequency and pulse duration from the interrogator it automatically transmits coded pulses in reply and this response identifies the target as friendly on the ground radar p.p.i.

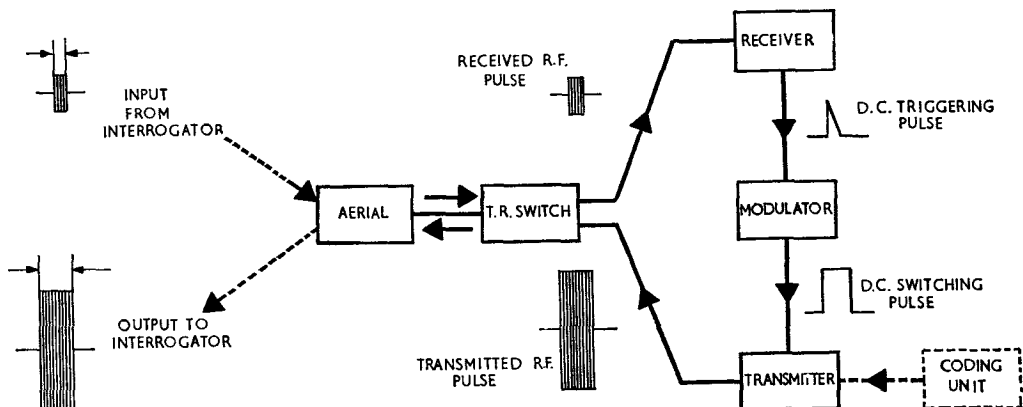


FIG 14. BLOCK SCHEMATIC DIAGRAM OF SECONDARY RADAR TRANSPONDER

Secondary radar may also be used as a navigational aid. In this role the interrogator is carried in the *aircraft* and the transponder is a radar beacon on the *ground*. The signals transmitted by the transponder when interrogated by the aircraft equipment are used by the aircraft to provide information on the range and bearing of the beacon. In this way a required landing ground may be located.

The operation is basically the same in each case. The block diagram of a typical system is shown in Fig 14. Each pulse from the interrogator is picked up by the transponder aerial and passed to the receiver where it is amplified and converted to a d.c. pulse. This pulse triggers the modulator which produces a waveform to switch the transponder transmitter on and off. A coding unit may be used to regulate the duration of the transponder pulse, this pulse being radiated back to the interrogator where it is used to provide the required information about the 'target'.

CHAPTER 3

FACTORS AFFECTING THE PERFORMANCE OF PULSE-MODULATED RADARS

Introduction

Although we now know something of the basic operation of a pulse-modulated primary radar there are still many questions left to be answered (Fig 1). In this chapter we shall attempt to answer these, and similar, questions.

WHAT DETERMINES MAXIMUM RANGE?
WHY SUCH HIGH POWERS?
WHY ARE HIGH FREQUENCIES NEEDED?
WHAT ARE THE LIMITATIONS OF RADAR?
EFFECTS OF UNWANTED OBJECTS?
DISCRIMINATION BETWEEN ADJACENT TARGETS?
WHAT IS THE RELATIONSHIP BETWEEN P.R.F.
AND PULSE DURATION?



FIG 1. QUESTIONS CONCERNING RADAR PERFORMANCE

Terms Used in Pulsed Radar

We have already met some of the terms used in pulse-modulated radar. Let us now define those terms more precisely and then go on to show their relationship with other, equally important, terms (Fig 2).

The waveforms shown in Fig 2 are 'ideal', i.e. we have assumed zero rise and decay times. In practice each pulse is more rounded because it takes a finite time to rise and to fall. However the ideal waveforms shown are adequate for our purpose. Note that the axes used in Fig 2 are power against time although more usually the waveform refers to *voltage* and time.

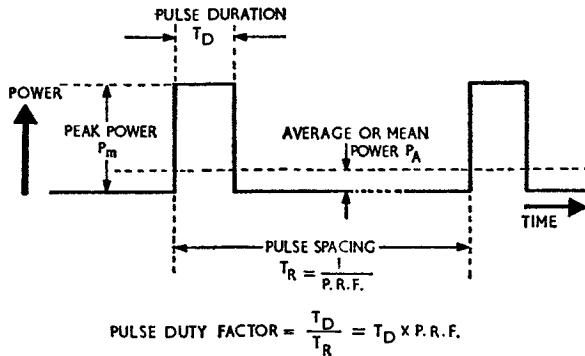


FIG 2. TERMS USED IN PULSED RADAR

Pulse repetition frequency. This is the *number* of pulses occurring in one second. For a p.r.f. of 500 p.p.s. the pulse spacing T_R is $\frac{1}{500}$ second or 2,000 microseconds. The value of p.r.f. normally lies between about 200 and 6,000 p.p.s.

Pulse duration. The pulse duration T_D is the *length of time* for which the transmitter is switched on to give each pulse. It normally lies between about 0.1 and 10 μs .

Pulse duty factor. This is also known as the *duty cycle*. It is the *ratio* of the pulse duration T_D to the pulse spacing T_R . Since $T_R = \frac{1}{\text{PRF}}$, the pulse duty factor is also the *product* of the pulse duration and the p.r.f. If the pulse duration is $1\mu\text{s}$ and the p.r.f. 500 p.p.s. the pulse duty factor is $10^{-6} \times 500$ or $\frac{1}{2,000}$, which is usually read as 1 in 2,000. It means that there is one $1\mu\text{s}$ pulse to every 2,000 μs . If we drew each pulse with a width of 1 inch to represent the pulse duration of $1\mu\text{s}$ the next pulse would occur 2,000 inches further along, i.e. approximately 55 yards away!

Peak and average values of power. We have seen that the power output of the transmitted pulses should be as great as possible; but we also know that when power is developed in a circuit *heat* is produced in the components. However, because the transmitter works in pulses we are able to develop a *very high power* in the circuit for the short duration of the pulse and, in the *comparatively long resting time* between pulses, the heat produced can be removed. If the transmitter uses a pulse duration of $1\mu\text{s}$ and a p.r.f. of 500 p.p.s. in *one hour* of operation the transmitter is switched on for a total time of only 1.8 seconds. Although the *peak power* developed during each pulse may be *very high* the *average* or *mean power* over a long period is quite *low*. The power rating of the components used in the transmitter circuit can therefore be much lower than might at first seem necessary.

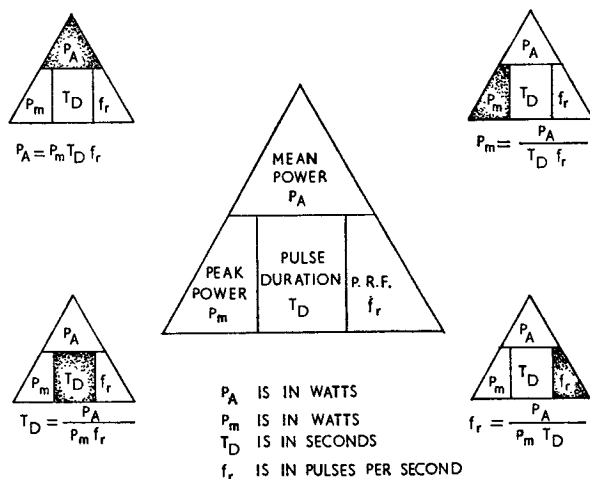


FIG 3. RELATIONSHIP BETWEEN PEAK POWER, PULSE DURATION, PRF AND MEAN POWER

Relationship between the various factors. There is a very simple relationship between the peak power during the pulses, the pulse duration, the p.r.f. and the mean power. From Fig 3 we see that the mean power is equal to the peak power multiplied by the product of pulse duration and p.r.f. Since pulse duration $\times$ p.r.f. equals the pulse duty factor the mean power P_A may be written as:

$$P_A = \text{Peak power } P_m \times \text{duty factor.}$$

Example. If a radar transmitter radiates a peak power of 1MW with a pulse duration of $1\mu\text{s}$ and a p.r.f. of 1,000 p.p.s. the mean power is:

$$\begin{aligned}
 P_A &= P_m \times T_D \times f_r \\
 &= 10^6 \times 10^{-6} \times 1,000 \\
 \therefore P_A &= 1 \text{ kW.}
 \end{aligned}$$

Thus a transmitter rated at 1 kW mean power can be used to produce 1 MW pulses if the pulse duty factor is $T_{Df_r} = \frac{1}{1,000}$.

Factors Affecting Radar Operation

Many factors affect the operation of a radar installation. Some of these factors are external to the radar set and affect all radars, e.g. reflections from unwanted objects and effects of external noise and jamming. We have very little control over such things.

Other factors controlling the performance of the radar are the result of deliberate decisions in design, e.g. the frequency of operation and the values of p.r.f. and pulse duration. These are selected and adjusted to give specific results and thus *intentionally* limit the operation of the radar.

External Factors Affecting Performance

The main external factors limiting the performance of a radar are:

- a. Sources of external noise.
- b. Reflections from unwanted objects.
- c. The dimensions of the target.

Let us examine each of these in more detail.

External Noise

We have previously seen the effects of noise on a c.r.t. display. External noise may be due to any of the factors illustrated in Fig 4. In some circumstances (c) may be the result of *deliberate* 'jamming' of the radar by an enemy transmitter; this jamming technique is called "Electronic Counter Measures" (ECM) and uses a specialized radar.

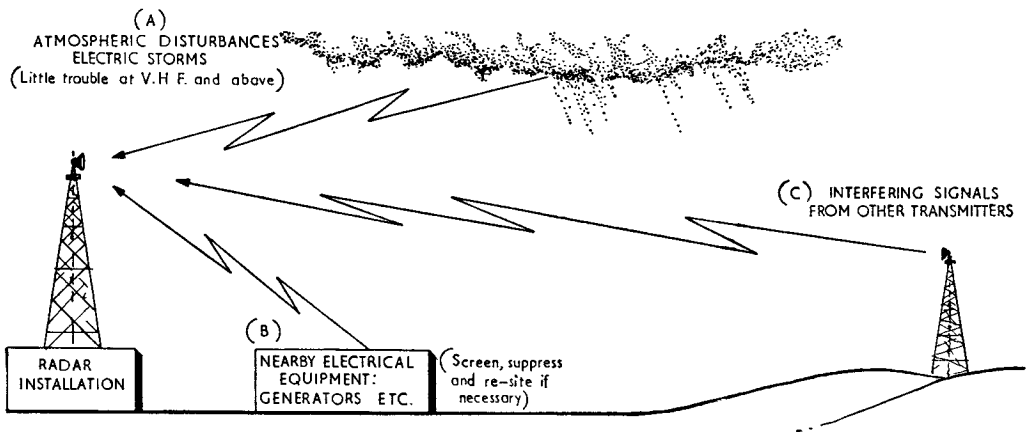


FIG 4. SOURCES OF EXTERNAL NOISE

Reflections From Unwanted Objects

We must first define what we mean by 'unwanted objects'! Something which may be an unwanted echo to one type of radar may be the wanted echo to another type. For example, clouds, built-up areas, hills and aircraft all produce radar echoes. For an early-warning search radar the wanted echoes are those produced by aircraft or missiles. For meteorological radar, on the other hand, it is the cloud formations which produce the wanted echoes. For bombing radar the towns and hills produce the wanted echoes, i.e. the 'radar map'.

In early-warning search radar, echoes produced by anything other than aircraft or missiles are unwanted. We are then faced with the problem of removing the 'clutter' on the c.r.t. display due to these unwanted echoes (Fig 5). Such clutter may be so pronounced, especially from objects near the ground radar, as to completely hide the wanted echoes. Fortunately modern radar contains circuits which can differentiate between moving objects and stationary objects. By using *moving target indication* radars (MTI) the appearance of stationary objects on the display may be suppressed so that the majority of the clutter is removed. Such systems will be considered in more detail later.

Dimensions of Targets

All pulse-modulated primary radars depend for their success upon the reflection of e.m. energy from objects which have been 'illuminated' by a radar beam. All objects in the path of the beam will reflect energy to some extent. The amount reflected depends upon the *material* of which the object is made, the *shape* of the object and its *size* (Fig 6).

If we have two identical objects at different distances from the radar the one nearer the radar reflects more energy.

A metal object will reflect more energy than an object of the same size and shape made of wood or plastic. The *better* the conductor the *greater* is the reflection.

The *shape* of the object will determine *how* the energy is reflected. If the object has a flat side facing the radar transmitter it will reflect more of the energy *back towards* the radar than an object of any other shape.

Large objects will reflect more energy than small objects of the same material and shape at the same distance from the transmitter. The object however must be greater than a certain *minimum* size, in terms of wavelength of the radiated energy, to produce a reasonable reflection of energy. Generally targets must have a size *greater* than about a

quarter of the radar wavelength being used before a detectable echo is received. Thus for the detection of *small* objects the radar wavelength must also be small, i.e. the frequency must be *very high*. This is one reason for the use of high frequencies in radar.

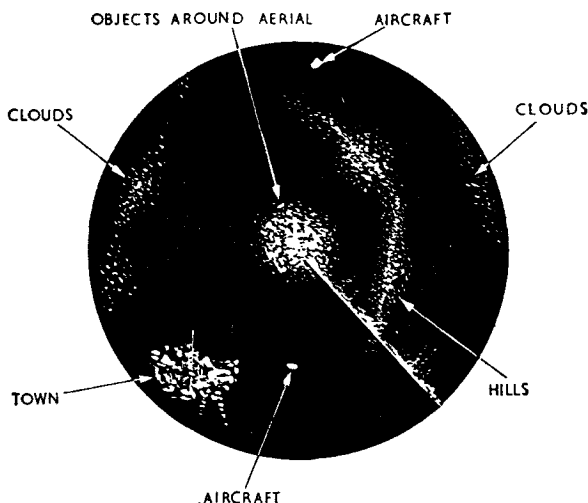


FIG 5. CLUTTER ON A PPI DISPLAY

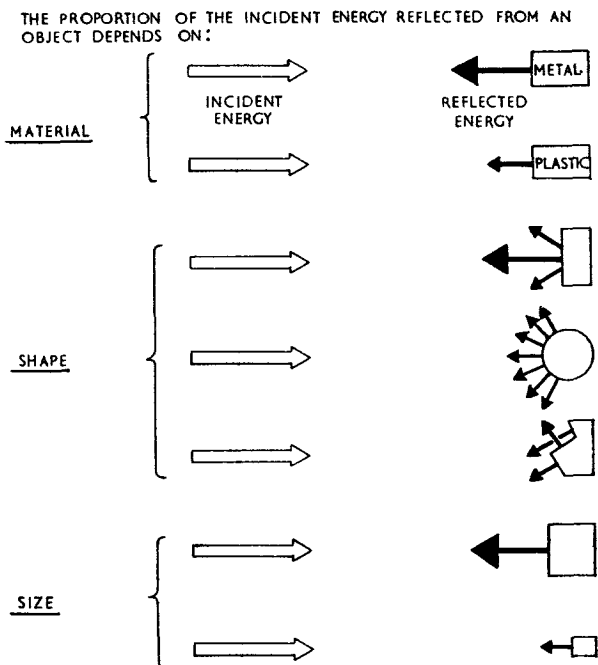


FIG 6. CHARACTERISTICS OF TARGETS

Design Factors Affecting Radar Performance

The main factors in the design of a radar set which affect the performance of the radar are:

- Transmitter power.
- Receiver sensitivity and noise factor.
- Frequency of operation.
- Shape of radar beam and scanning methods used.
- Pulse repetition frequency.
- Pulse duration.

Let us now examine each of these factors in more detail.

Transmitter Power

Even with the most concentrated radar beam only a *fraction* of the energy of each radiated pulse strikes the target. At the target this fraction of the original energy is 'scattered' so that, in turn, only a *fraction* of the *incident* energy returns towards the receiving aerial. To compensate for this very inefficient reflecting process the greatest possible radiated power must be used. This is why we use peak powers of 1MW and more; but even with such high powers the power in the received echo is only of the order of milliwatts or even microwatts.

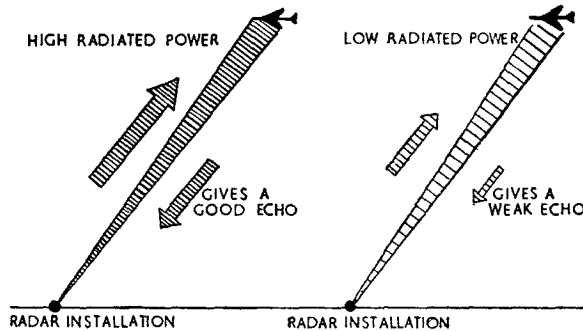


FIG 7. EFFECT OF TRANSMITTER POWER OUTPUT ON RECEIVED SIGNALS

In general the *higher* the radiated power the *greater* is the received echo power and hence the *greater* is the range (Fig 7). However the increase in range obtained by increasing the radiated power is very small. Even *doubling* the power increases the range only *1.19 times*.

The power that the transmitter is designed to radiate depends upon the job that the radar has to do. Obviously a long-range search radar needs a very high peak power during the pulses. The power needed by an airfield approach radar, where the required range may be only a few miles, is very much less.

Receiver Sensitivity and Noise Factor

We have already seen that the main limitation on useful amplification in a radar receiver is the relationship between the amplitude of the wanted signal voltage and that of the noise voltage, i.e. the *signal-to-noise ratio*. If the input has a low signal-to-noise ratio the signal echo on the c.r.t. may be 'lost' among the noise indications. The input signal-to-noise ratio to a receiver is determined by external factors as previously noted and it is, at the moment, the ultimate limitation on the reception of very weak echoes.

In addition, the receiver itself 'generates' noise (valve noise and thermal agitation) and the receiver noise, when combined with the input noise, means that the *output* signal-to-noise ratio is lower than the *input* signal-to-noise ratio. The *ratio* of the signal-to-noise ratio at the input to

that at the output is known as the 'noise factor' of a receiver. It is a measure of the noise introduced by the receiver itself (Fig 8). The design problem is to produce a receiver with as low a noise factor as possible. This is complicated by the fact that the receiver has to accept very narrow pulses and hence a *wide band* of frequencies. Wide bandwidth tends to *increase* the noise factor of a receiver. Thus the design of a receiver is a compromise between high sensitivity, wide bandwidth and low noise factor.

Frequency of Operation

The frequencies used in radar are high for three main reasons (see Fig 9):

- To obtain a good echo the radar wavelength must be *less than four times* the size of the target.
- For good *angular discrimination* between adjacent targets, for accurate indication of bearing and for adequate concentration of the radiated energy we use aerials which can provide a *very narrow beam*. This can be achieved much more easily at high frequencies.
- High frequencies are needed to ensure an adequate number of r.f. cycles in each pulse.

The frequency chosen for a particular radar depends upon the job it has to do. A high-resolution radar which is required to discriminate between targets very close together in bearing will use super-high frequencies in the microwave region—in the band 3,000 to 30,000 Mc/s. For long range radars, where early warning is the criterion and accuracy of range and bearing of less importance, the v.h.f. band around 200 Mc/s may be used. The lower frequencies have the advantage of smaller atmospheric absorption and longer ranges.

In radar the *wavelength* at which the equipment is operating is quoted as often as the frequency. The relationship between frequency f in cycles per second, velocity c of e.m. waves in metres per second and wavelength λ in metres is illustrated in Fig 10.

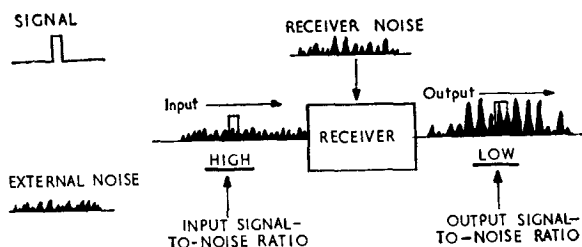


FIG 8. RECEIVER NOISE FACTOR

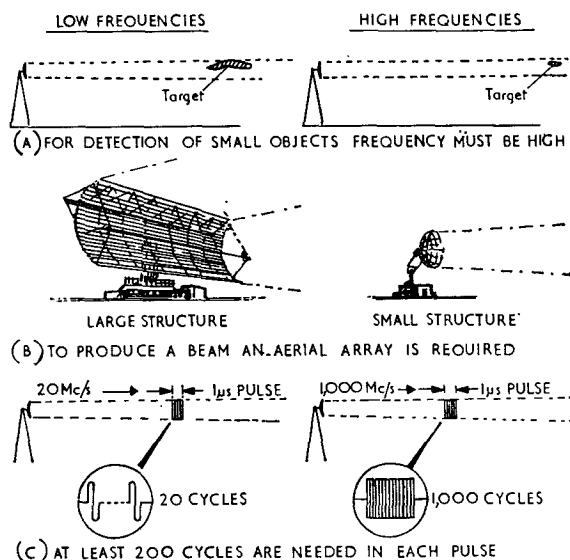


FIG 9. NEED FOR HIGH FREQUENCIES IN RADAR

VELOCITY OF E.M. WAVES C = WAVELENGTH λ X FREQUENCY f

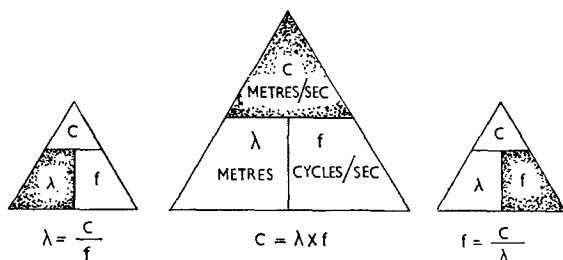


FIG 10. RELATIONSHIP BETWEEN FREQUENCY AND WAVELENGTH

The most common frequencies used in radar, together with their corresponding wavelengths, are in bands as shown in Table 1.

Band	Centred on— Mc/s	Centred on— cm	Remarks
P	300	100	Metric
L	1,200	25	Centimetric
S	3,000	10	"
C	5,000	6	"
X	10,000	3	"
J	13,000	2.25	"
K	24,000	1.25	"
Q	40,000	0.75	Millimetric
V	60,000	0.5	"
O	100,000	0.3	"

TABLE 1—RADAR BANDS

There are considerable differences in the design of radars for use in the metric, centimetric and millimetric regions. These differences are mainly in circuitry, components and techniques and will be discussed in later chapters.

Shape of Radar Beam and Scanning Methods

We have already noted the need for narrow beams. The *shape* of the beam however will depend upon the requirement of the radar. The *pencil beam*, the shape we have been concerned with till now, gives a high precision of angle measurement, but the *time* taken to scan a given volume of sky may be too long for quick detection and tracking of a target. Very often other beam shapes and other scanning methods are used to determine *quickly* the required information about a target.

One quick method of scanning involves the full rotation of a *fan beam* about a vertical axis (Fig 11). Using a beam like this in conjunction with a p.p.i. gives accurate determination of range and bearing *in azimuth* once every revolution of the aerial. *No indication* of elevation is possible.

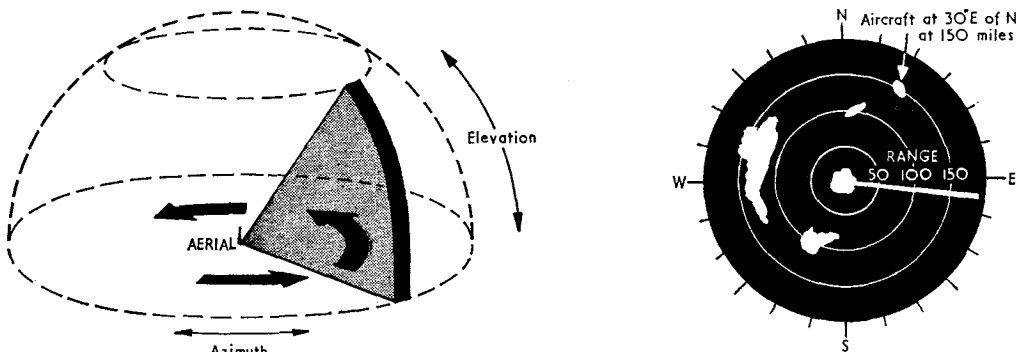


FIG 11. USE OF A FAN BEAM

Fig 12 illustrates a *beaver-tail beam*. This beam, which is narrow in the vertical (elevation) plane and broad in the horizontal (azimuth) plane, is 'nodded' up and down between the scanning limits. It may be used in conjunction with a type B display or a range-height indicator to give quick and fairly accurate determination of the *elevation* of the target.

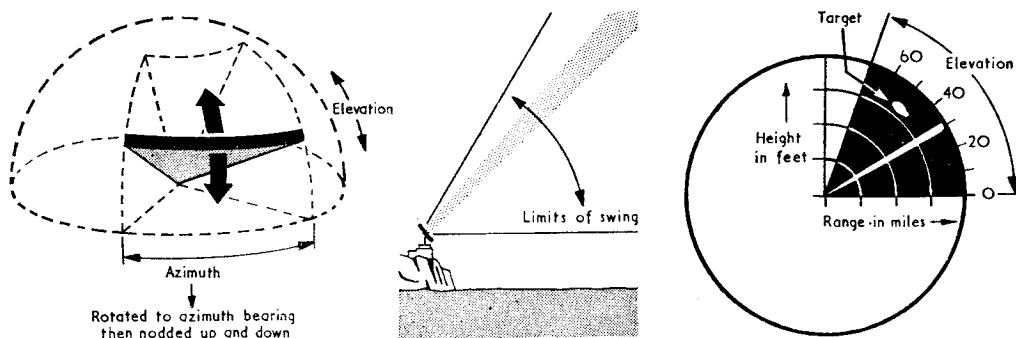


FIG 12. USE OF A BEAVER-TAIL BEAM

Some ground radars use a fan beam and a beaver-tail beam in conjunction to give accurate bearings in azimuth and elevation respectively. This requires two separate aerials, one for each beam shape.

Some radars (especially those carried in aircraft) cannot afford the luxury of two separate aerials. In such cases a single aerial, producing a *pencil beam*, is caused to rotate a full 360° in azimuth, a relatively slow upward tilt being imposed on the aerial at the same time to give

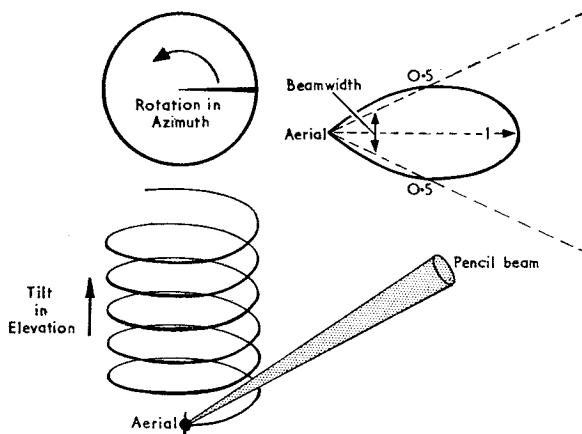


FIG 13. HELICAL SCANNING

coverage in elevation. This is known as *helical scanning* (Fig 13). One type of radar aerial using helical scanning has a beam width of 8° , a tilt of 70° and rotates at 180 r.p.m.; a complete scan takes about six seconds.

Note from Fig 13 that the *beam width* is the angle between points where the radiated power has fallen to half its maximum value (the 'half-power' or 3db points).

In *conical scanning* (Fig 14) a pencil beam is used and the axis of the beam rotates to sweep out a circular cone. This type of scan may be used after the approximate bearings in azimuth and elevation have been obtained by other means, e.g. by helical scanning. When the axis of the

cone is in line with the target (Fig 14a) the aerial is correctly 'aimed'. If the aerial is incorrectly aimed (Fig 14b) the signals received from the A and B positions are proportional to OA and OB

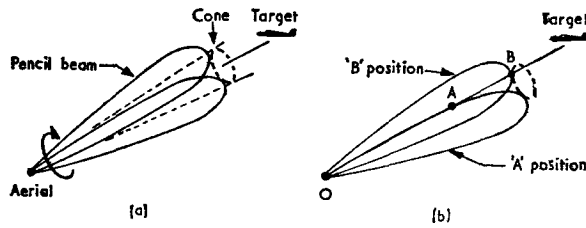


FIG 14. CONICAL SCANNING

respectively. By comparing the two signals the magnitude and sense of the error can be calculated and a correction applied to give automatic 'following' or 'tracking' of an object.

Spiral scanning is a development of the conical scan. A pencil beam is produced by the aerial and the beam is then made to rotate about a *horizontal* axis in ever-increasing circles to produce a spiral as shown in Fig 15. Angles up to about 60° in both azimuth and elevation may be scanned in this way.

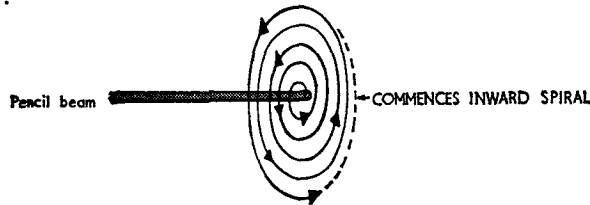


FIG 15. SPIRAL SCANNING

In all the systems so far considered the scan is the result of *physical movement* of the aerial. In large ground radars, where the aerial assembly is very wide, it is inconvenient to swing the aerial because of the large forces involved and because of the space taken up by a large rotating aerial. In such cases the *beam* is caused to scan, *without physical movement of the aerial*, by varying the phase of the input currents to the aerial elements in a cyclic manner. These *electronic scanning methods* are efficient and economical and will be considered in more detail later.

Pulse Repetition Frequency

The p.r.f. selected for a particular radar depends upon several factors, the most important being summarized below:

a. Maximum required range. Each pulse must be given time to travel to the most distant required target and return before the next pulse is transmitted otherwise there will be a risk of confusion on the display. If the maximum required range of a radar is 100 miles, the time taken for a pulse to travel to a target at 100 miles range and return is $100 \times 10 \cdot 75 = 1075 \mu\text{s}$. The transmitter must be quiescent for at least this time, i.e. the pulse spacing T_R must have a *minimum* value of $1075 \mu\text{s}$. Since $\text{p.r.f.} = \frac{1}{T_R}$ the *maximum* value of p.r.f. is

$$\frac{10^6}{1075} = 930 \text{ p.p.s.}$$

If the p.r.f. is adjusted to this maximum value then any signals received from targets at a range *greater* than 100 miles would appear in the *next* pulse spacing period. Thus in Fig 16 target A, within the required range, is indicated on the c.r.t. during *every* pulse spacing period. Target B, outside the required range, *misses the first pulse-spacing period*. The *persistence*

of the c.r.t. screen is normally sufficient to retain the indication of the target B all the time and, to the operator, target B appears to be at a range of 40 miles (Fig 16).

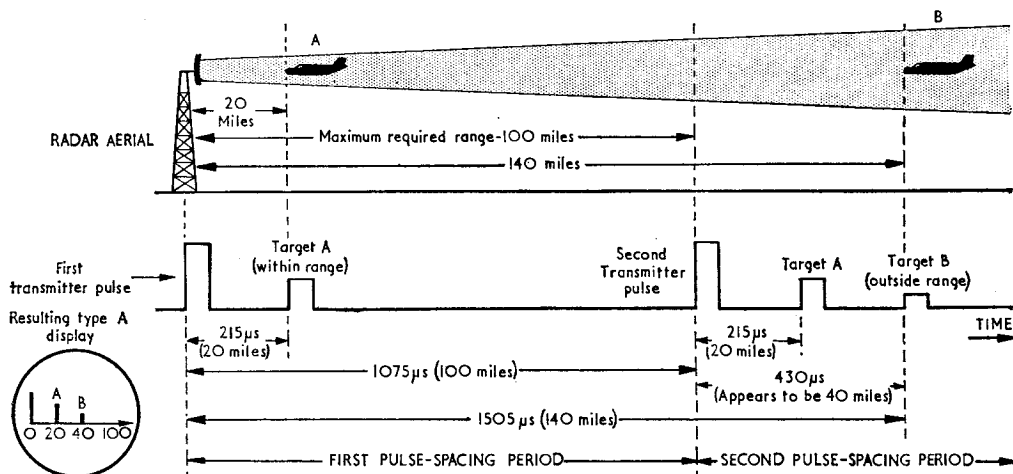


FIG 16. INDICATIONS FROM TARGETS BEYOND THE REQUIRED RANGE

To avoid this, the pulse spacing is made *very much* longer than its minimum value of 1075 μs, i.e. the p.r.f. is *reduced well below* 930 p.p.s. The c.r.t. timebase trace however is still adjusted for a sweep time of 1075 μs. If a p.r.f. of 200 p.p.s. is selected the result will be as shown in Fig. 17. Any target beyond 100 miles range will return an echo *after* 1075 μs, when the timebase trace has been switched off, and so will produce *no indication* on the c.r.t. screen.

The same reasoning applies for any required radar range, but for *smaller* ranges the time intervals are *less* and the p.r.f. may be *increased*. Thus the *longer* the required range the *lower* must be the p.r.f.

b. Scanning speed. If the aerial scanning speed is *high* and the p.r.f. is *low* some targets may be missed because, in the time interval between pulses, the aerial will have turned through a certain angle. Thus the *higher* the scanning speed the *higher* must be the p.r.f. Usually however the p.r.f. is decided by other factors and it is then the scanning speed which is adjusted to suit the selected p.r.f. The speed of scan is related to the p.r.f. by the expression $f_s \theta$, where f_s is the p.r.f. in p.p.s. and θ is the half-power beam width in degrees. For a beam width of 1° and a p.r.f. of 500 p.p.s. the scanning speed must be less than $500 \times 1^\circ = 500^\circ$ per second or 1.4 r.p.s. (84 r.p.m.). This ensures that at least one pulse per beam width is transmitted.

c. Mean power available. Given a certain available mean power and a required pulse duration, the p.r.f. may have to be adjusted to produce an acceptable peak power output. This may be seen from the relationship given earlier:

$$\text{Peak power} = \frac{\text{Mean power}}{\text{PRF} \times \text{Pulse duration.}}$$

If the mean power available is 1 kW and a pulse duration of 5 μs is required then, with a p.r.f. of 2,000 p.p.s., the peak power during each pulse is $\frac{10^3}{2 \times 10^3 \times 5 \times 10^{-6}} = 100 \text{ kW}$. If we require a *larger* peak power for the same values of mean power and pulse duration the

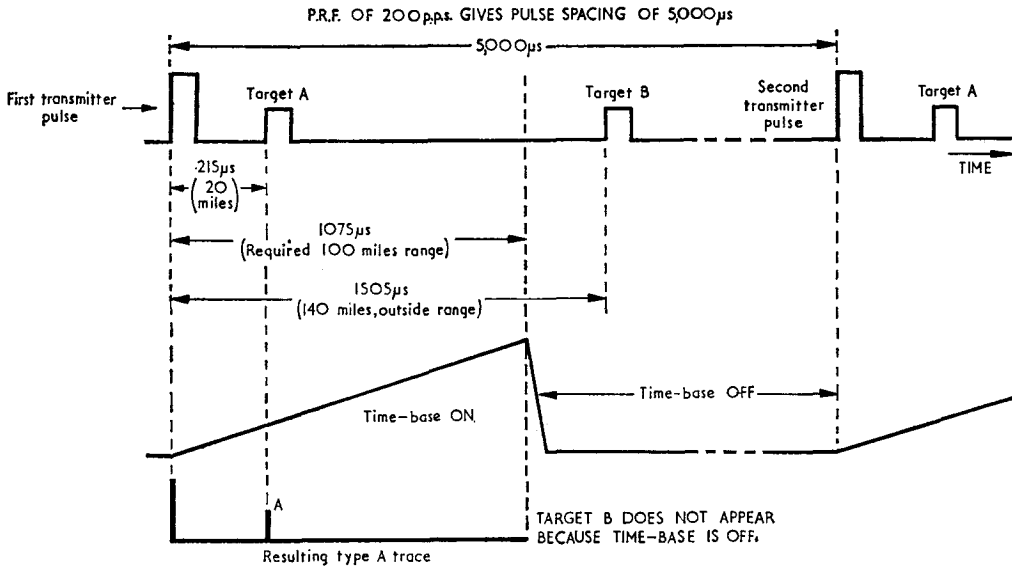


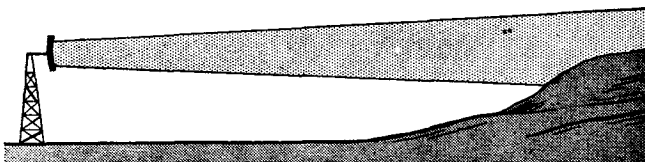
FIG 17. ADJUSTMENT OF PRF AND TIME-BASE TO LIMIT INDICATIONS TO REQUIRED RANGE

p.r.f. must be *reduced*. For a p.r.f. of 200 p.p.s. the peak power output is now

$$\frac{10^3}{200 \times 5 \times 10^{-6}} = 1 \text{ MW.}$$
 What we are really doing here is to adjust the pulse duty factor—the ratio of pulse duration to pulse spacing.

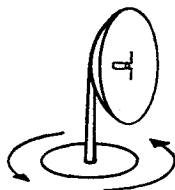
d. Improved definition. The more pulses that are transmitted per second the more pictures are 'painted' on the c.r.t. per second. This means a brighter and clearer display and improved definition. In addition the *higher* the p.r.f. the more echoes that are received from any one target per second. This means more reports per second on the *change* in range and direction of a moving target.

From what has been said it is clear that the p.r.f. selected for a particular radar must be a compromise between several conflicting requirements (Fig 18). Because of this, many radars have more than one value of p.r.f. for different ranges.



(a) RANGE

LONG RANGE—LOW P.R.F.



(b) SCANNING SPEED

HIGH SCANNING SPEED—HIGH P.R.F.

FIG 18. (a-b) FACTORS DETERMINING PRF

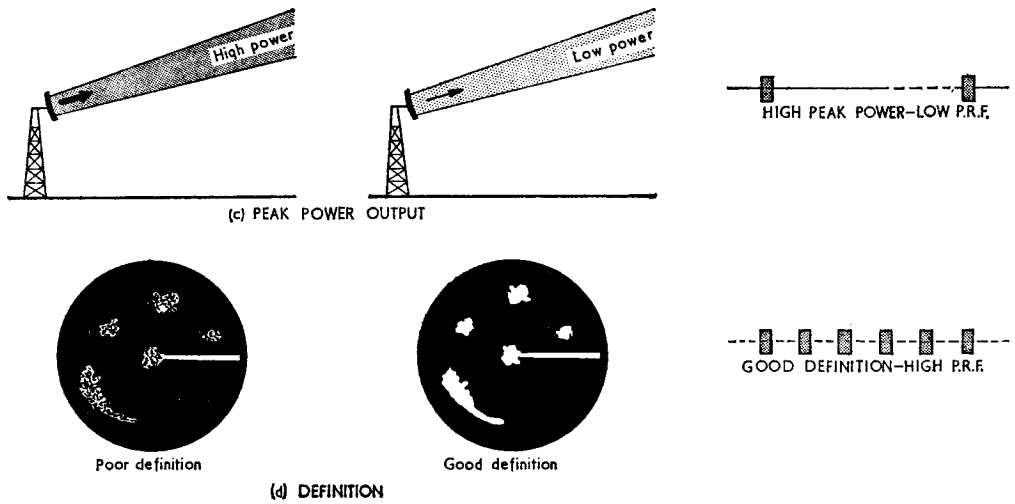


FIG 18. (c-d) FACTORS DETERMINING PRF

Pulse Duration

The pulse duration selected for a particular radar depends upon several factors, the most important being summarized below:

- a. **Minimum range.** The time interval equivalent to one radar mile is $10.75 \mu s$. Thus if we are using a pulse duration of $10.75 \mu s$ any target at a range *less* than one mile will not be seen because the transmitter will still be firing when the target echo arrives back at the aerial.

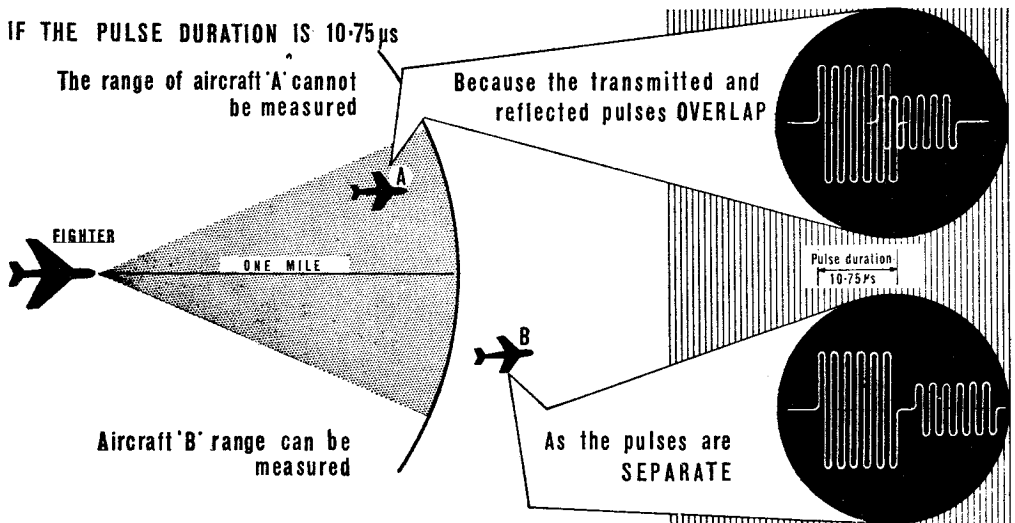


FIG 19. EFFECT OF USING LONG-DURATION PULSES FOR SHORT-RANGE TARGETS

Since the aerial is still connected by the T-R switch to the transmitter the echo will not provide any input to the receiver (Fig 19). *Short-duration* pulses are needed for *short-range* working.

There is an area surrounding any pulsed radar within which no targets can be detected. For long-range search radars this 'blind' area is relatively unimportant. However if the radar is an airborne interception (AI) equipment carried in a fighter aircraft it is important that target echoes are displayed on the radar right up to the interception point. This may be only a few hundred feet.

A pulse duration of $10.75 \mu\text{s}$ produces a blind area of one mile (5,280 feet). If we require interception down to 100 feet the pulse duration must not exceed $\frac{10.75 \times 100}{5,280} = 0.2 \mu\text{s}$. This is a typical value of pulse duration for the low ranges of an airborne interception radar.

b. Target discrimination in range. Since e.m. waves travel 300×10^6 metres in one second, a pulse of $1 \mu\text{s}$ will extend 300 metres along the direction of propagation. If there are *two* targets *within* that 300 metres they will be simultaneously 'illuminated' by the pulse and may appear as a *single echo* at the receiver (Fig 20). *Short* pulse durations are needed for good

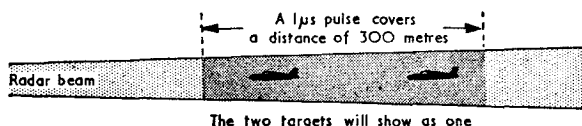


FIG 20. RANGE DISCRIMINATION OF TARGETS

target discrimination *in range* in the same way that very narrow beams are needed for good discrimination *in angle*. To distinguish between two targets on the same bearing and elevation, but separated from each other in range by 100 feet, a pulse duration of less than $0.2 \mu\text{s}$ is needed (using the figures of the previous paragraph).

c. Frequency used. The *lower* the frequency the *longer* must be the pulse duration to ensure an adequate number of r.f. cycles in each pulse.

d. Mean power available. Given a certain available mean power and a selected p.r.f. the pulse duration may have to be adjusted to produce an acceptable peak power output. This may be seen from the relationship:

$$\text{Peak power} = \frac{\text{Mean power}}{\text{PRF} \times \text{Pulse Duration}}.$$

If the mean power available is 1 kW and the p.r.f. is 5,000 p.p.s. a pulse duration of $2 \mu\text{s}$ will produce a peak power output of $\frac{10^3}{5 \times 10^3 \times 2 \times 10^{-6}} = 100 \text{ kW}$. By reducing the pulse duration to $1 \mu\text{s}$ (the other factors remaining constant) the peak power output increases to $\frac{10^3}{5 \times 10^3 \times 10^{-6}} = 200 \text{ kW}$. Again, what we are really doing is adjusting the pulse duty factor.

e. Receiver bandwidth. We have already noted that the *shorter* the pulse duration the *greater* is the band of frequencies associated with the pulse and the greater must be the *bandwidth* of the receiver accepting such pulses. If the pulse duration is too short for the available receiver bandwidth, distortion and attenuation of the pulse result (Fig 21). This leads to inaccuracies in range measurement. Thus where the receiver bandwidth has been limited for other reasons (e.g. reduction of noise) the pulse duration must not be made *too short*.

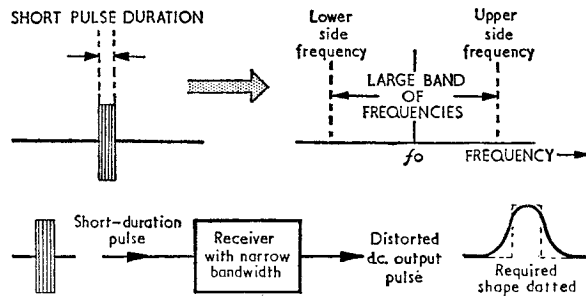


FIG 21. PULSE DURATION AND RECEIVER BANDWIDTH

Again, like the p.r.f., the value of pulse duration selected for a radar is a compromise between conflicting factors (Fig 22). Because of this some radars have several values of pulse duration which may be selected as required.

Some Typical Applications

We can see from what has been said that a radar set is a compromise. Great range is incompatible with good range resolution, high accuracy and high scanning speeds. The selection of the p.r.f. and pulse duration depend upon many factors, some requiring high values and others low values. Steep and narrow pulses are necessary but this means increasing the receiver bandwidth and hence the noise factor. If the bandwidth is reduced the pulse is distorted. All these conflicting requirements are considered by the designer. The result is that a radar set designed for use in one role would be of little value in another role. Each application requires different variations in all the variable factors.

It would be impossible to consider more than a very small number of typical applications

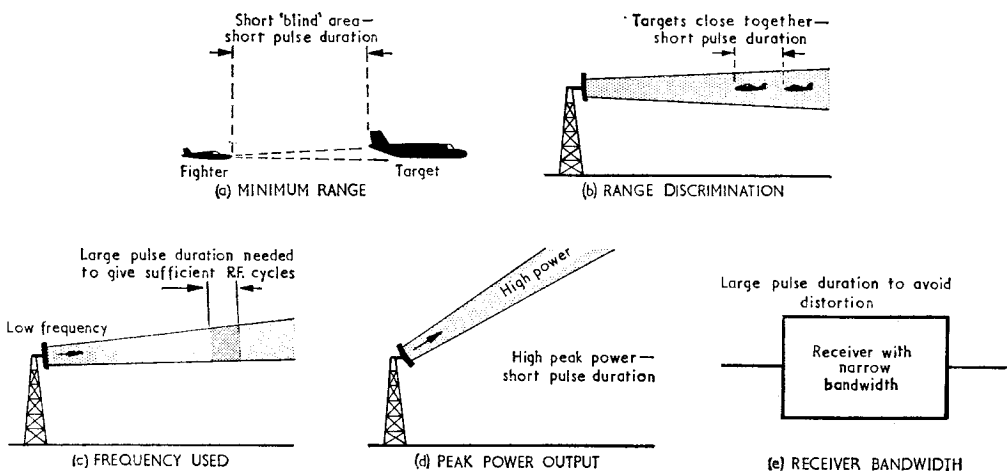


FIG 22. FACTORS DETERMINING PULSE DURATION

in a book of this type. Table 2 lists the main characteristics of the seven different radars noted below:

- a. A *metric* long-range early-warning equipment (radar type 7).
- b. A *centimetric* long-range early-warning equipment (radar type 80).
- c. The *search* portion of a radar set (radar type CPN-4) designed to locate an aircraft at a range of about 20 miles so that by verbal instructions over radio it may be guided to an airfield during periods of poor visibility.
- d. The *precision* portion of the same set (CPN-4) designed to guide the aircraft from about six miles away to within about a half mile of the runway.
- e. An *aircraft bombing and navigation aid* (NBS Mk 1A).
- f. An *airborne interception* equipment (AI Mk 23).
- g. A *secondary radar interrogator* equipment (Rebecca Mk. 8).

Equipment	Frequency	PRF (p.p.s.)	Pulse Duration	Peak Power	Range (miles)
Radar type 7	220 Mc/s	250	8 μ s	450 kW	250
Radar type 80	3,000 Mc/s	270	2 or 5 μ s	1MW	300
CPN-4 Search	3,000 Mc/s	1,500	0.5 μ s	0.5MW	30
CPN-4 Precision	10,000 Mc/s	5,500	0.18 μ s	45 kW	10
NBS Mk 1A	10,000 Mc/s	800	0.5 μ s		
		400	1 μ s	175 kW	—
		200	2 μ s		
AI Mk 23	8,500 Mc/s	1,000	1 μ s	175 kW	25
Rebecca Mk 8	200 Mc/s	180	3 μ s	1 kW	200

TABLE 2—IMPORTANT CHARACTERISTICS OF TYPICAL RADARS

CHAPTER 4

SOME EXAMPLES OF THE USES OF PULSED RADAR

Introduction

One of the easiest ways of consolidating the information given in the previous chapters is to consider some typical examples of the applications of pulse-modulated radar. Only a few examples are chosen because radar has very many applications. Of the examples selected only the basic outline and general idea of the operation of each can be considered at this stage. More detailed information is given in later chapters. Many of the equipments mentioned in the following paragraphs have been superseded by more modern types of radar, some of which are classified. Nevertheless it is only the basic principles and ideas with which we are concerned here and these remain unaltered.

Ground-controlled Interception Radar (GCI)

Let us see how radar may be used to assist in the interception and destruction of an incoming hostile aircraft. Fig 1 shows that there are several stages in a ground-controlled interception:

- a. The *detection* of the target at long range (300 miles or so) and the alerting of our defences. This is the function of the long-range early-warning search radars.
- b. More precise radars on the station are then brought into operation: one accurately determines the range and bearing in azimuth of the target; another gives accurate determination of range and elevation; and others may be required to give more complete coverage of targets at very low altitude.

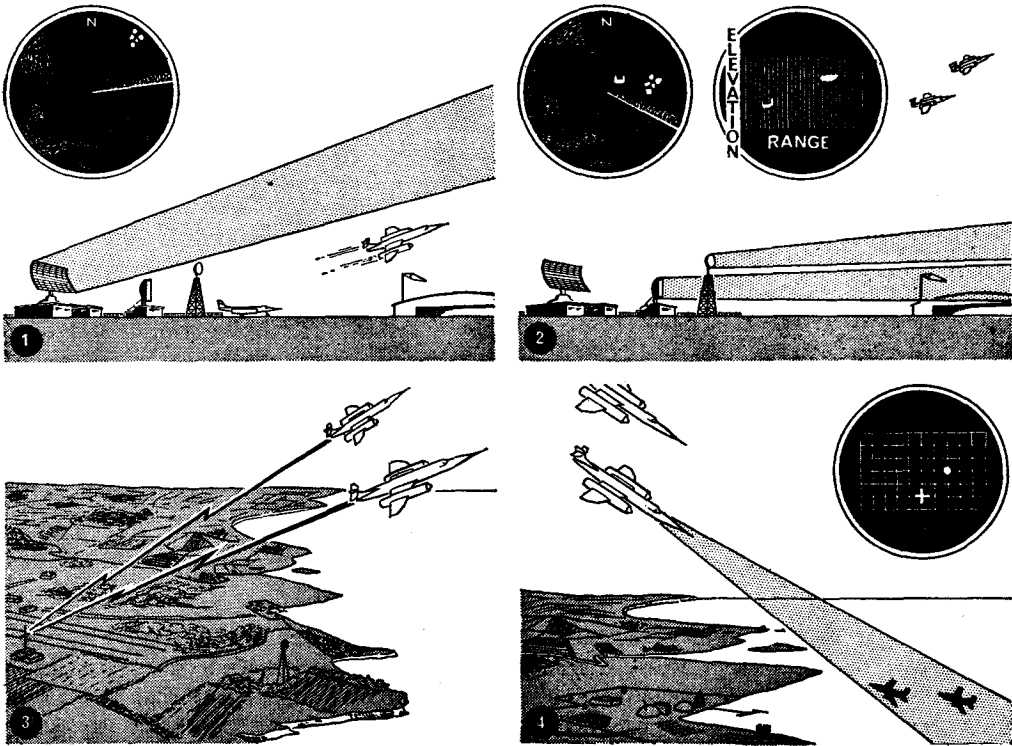


FIG 1. STAGES IN A GROUND-CONTROLLED INTERCEPTION

c. The information so obtained is passed to fighters (already airborne) or to an alerted surface-to-air missile station (SAM). If we consider the fighters, they will also show echoes on the ground display; but because they are carrying a secondary radar transponder known as i.f.f. (identification friend or foe) the echoes produced by the fighters are different from those produced by enemy aircraft. The fighters may then be directed towards the target by orders from the ground radar controller.

d. When the fighters are within range their own airborne interception radar (AI) takes over to complete the interception.

Long-range Early-warning Search Radars

The most important function of a long-range search radar is the *early detection* of the target. For long ranges low values of p.r.f. and high values of peak power are necessary and, for good discrimination in range, short duration pulses are needed. In addition a quick rate of scan is called for to give complete coverage as quickly as possible. Because of the low p.r.f. this is most easily achieved by using specially shaped beams. The usual practice is to use a form of fan beam so that approximate bearings in *azimuth* are obtained fairly quickly. The elevation angles can be measured later. Typical of the long-range early-warning search radars are the radar type 7 and the radar type 80.

The radar type 7 is a *metric* radar operating at about 200 Mc/s. It has a peak power output of 450 kW and uses a pulse duration of $8 \mu\text{s}$ and a p.r.f. of 250 p.p.s. The beam produced by the aerial is narrow in the azimuth plane and has a wider lobe or lobes in the elevation plane. The bearing in azimuth is obtained by rotating the aerial through 360° about a central pivot (Fig 2a). A p.p.i. used in conjunction with this movement indicates range and bearing of the target.

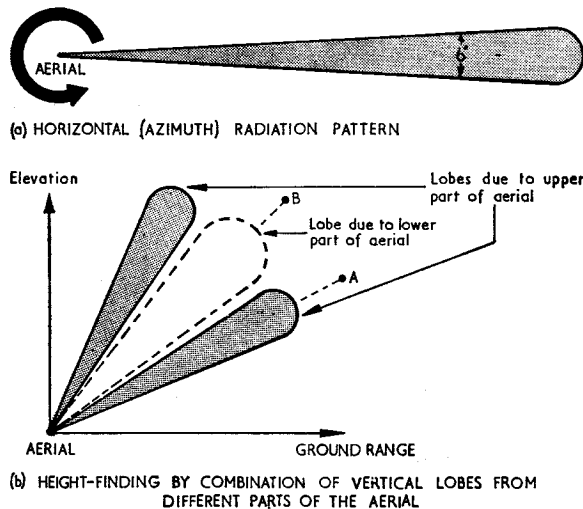


FIG 2. METRIC EARLY-WARNING RADAR, RADIATION PATTERNS

Approximate height-finding information can also be supplied by 'switching in' various vertical parts of the aerial to provide different lobe patterns in the elevation plane (Fig 2b). An aircraft at position A will produce a greater signal in the top part of the aerial than in the bottom part. An aircraft at position B will produce a greater signal in the lower part of the aerial than in the upper part. By comparing the signals received in the two parts of the aerial a measure of the angle of elevation of the target may be obtained.

The radar type 80 is a *centimetric* S-band radar operating at about 3,000 Mc/s. It has a peak power output of 2.5 MW and uses a pulse duration of $2\ \mu\text{s}$ or $5\ \mu\text{s}$ (depending upon the range discrimination required) at a p.r.f. of 250 p.p.s. The beam produced by the aerial is very narrow in the azimuth plane (about 0.3°). The bearing in azimuth, obtained by rotating the aerial through 360° (Fig 3a), is indicated on a p.p.i. The shape of the beam in the elevation plane is a modified fan beam (Fig 3b). This beam shape ensures that an aircraft at a given height produces a *constant amplitude* of echo for a large variation of elevation angle. A beam shape of this type is called a '*cosec squared*' pattern. Because of its shape *no indication* of the elevation or the height of the target is possible.

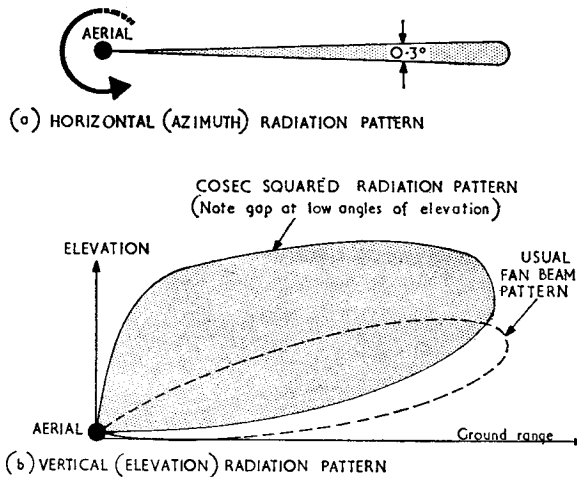


FIG 3. CENTIMETRIC EARLY-WARNING RADAR, RADIATION PATTERN

Medium-range Search and Height-finding Radars

As soon as approaching enemy aircraft are detected by the early-warning radars, fighters take off to intercept them and are guided to an attacking position by orders over a radio telephone link from the ground radar controller. To enable satisfactory instructions to be given, the ground radar must supply the controller with all the necessary information on range, bearing, height, track and speed of both the enemy and the friendly aircraft.

The equipment used for this is similar in many ways to the centimetric early-warning radar but a few special features influence its design:

- a. By the time the fighters have taken off and moved towards the point of interception the enemy aircraft will be much *closer* to its target. Thus the long-range requirements are less than those of the early-warning radars. Because of this the peak power output during the pulses need not be so high and the value selected for the p.r.f. may be *increased*.
- b. With a higher value of p.r.f. the definition of the display and also the rate at which information is supplied are both improved. In addition, the scanning speed may be increased. All these factors mean that *more accurate* assessment of the movement of the target can be made.
- c. In most cases the *approximate* values of range and bearing will have been obtained from the long-range early-warning radars. Thus the medium-range equipments may be concerned only with a *given volume* of space within which the target will be found. The procedure known as *sector scanning* can therefore be used (Fig 4).

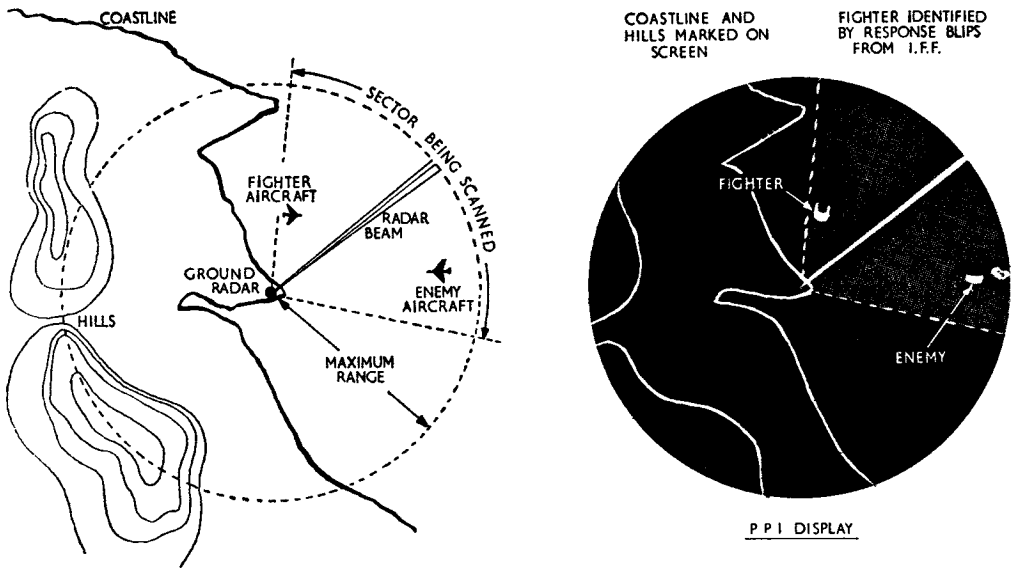


FIG 4. SECTOR SCANNING

Typical of the radars used to provide accurate information on all the relevant factors are the radars type 13 and type 14. The first provides accurate information on range and height and the other gives azimuth indication at small angles of elevation.

The radar type 13 is a centimetric S-band radar operating at about 3,000 Mc/s. It is a height-finding radar producing a peak power output of 0.5 MW with a pulse duration of up to $2 \mu s$ and a p.r.f. of up to 550 p.p.s. The aerial produces a *beaver-tail beam* whose width in the elevation plane is 1° and in the azimuth plane 5° (Fig 5a). The aerial is turned to the required azimuth bearing (obtained from the early-warning p.p.i.) and is then used for finding the elevation of the target on that bearing. This is done by 'nodding' the beam in the vertical plane between

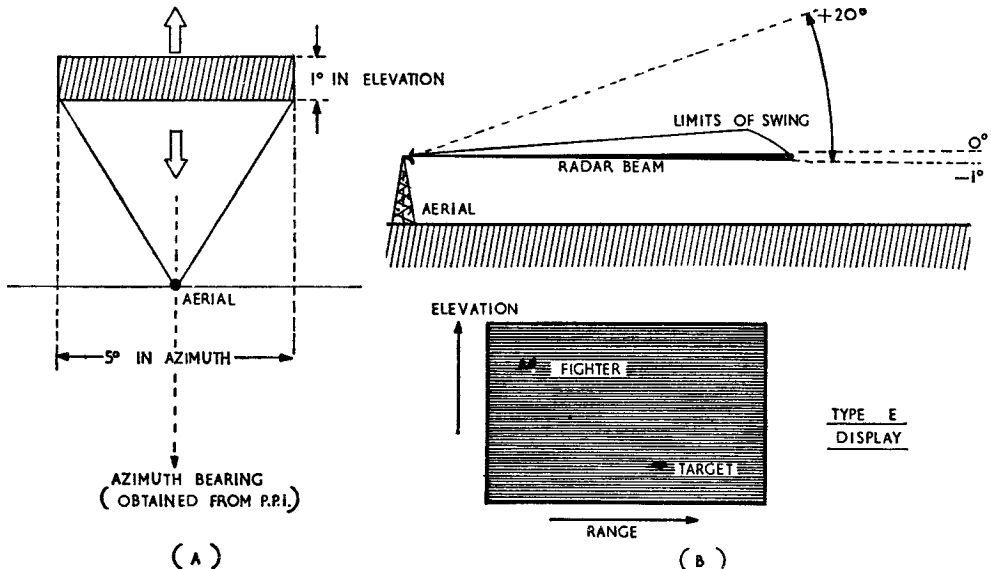


FIG 5. USE OF BEAVER-TAIL BEAM FOR HEIGHT-FINDING

-1° and $+20^\circ$ of elevation (Fig 5b). The display used is either a range-height indicator or a type E display.

The radar type 14 is similar to the radar type 13. The main difference is that it uses a different aerial to produce a *fan beam* (Fig 6). The beam width in the azimuth plane is now 1° and in the elevation plane it is about 4° . The aerial may be continuously rotated through 360° in azimuth, it may be aligned on any given heading or it may be used for sector scanning about a selected heading. A p.p.i. is used for the display. The aerial is normally pre-set to a tilt of $+1^\circ$ from the horizontal to give coverage at small angles of elevation. For greater coverage in elevation the aerial may be constructed to give a cosec-squared pattern (see earlier). Coverage of low-flying aircraft is then sacrificed.

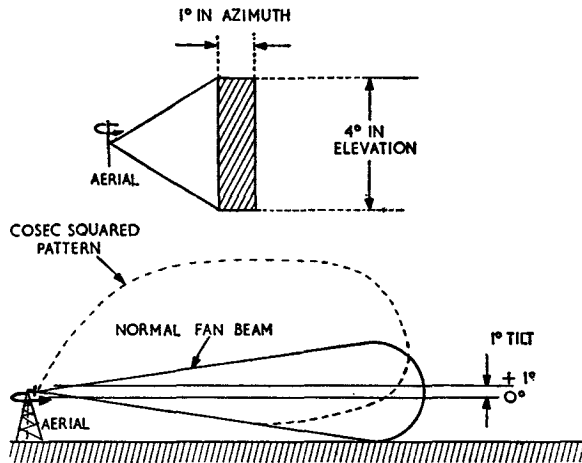


FIG 6. USE OF FAN BEAM IN PLAN SEARCH

Specifically designed to give *low-level* plan position indication of targets is the radar type 54. It is a centimetric S-band radar operating about 3,000 Mc/s. It has a peak power output of 0.5 MW at a p.r.f. of up to 550 p.p.s. and a pulse duration of up to $2 \mu\text{s}$. The aerial used with this radar is a *parabolic reflector* mounted on top of a high tower. It produces a *pencil beam* of approximately 2.5° beam width all round, at a low angle of elevation (Fig. 7). It may be used in the same way as the radar type 14 to give continuous coverage in azimuth or to scan a selected sector. In effect it 'fills the gap' of the radar type 14 coverage.

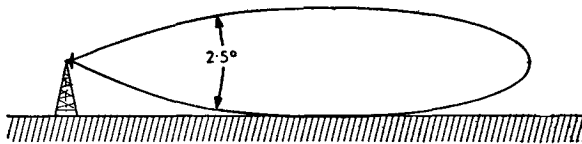


FIG 7. USE OF PENCIL BEAM IN PLAN SEARCH AT LOW ELEVATION ANGLES

By relating all the information displayed on the various p.p.i.s and height-finding displays the ground radar controller can guide the fighter into a favourable attacking position. The *direction* in which the enemy aircraft is moving can be seen by the movement of its echo on the displays and its *speed* may be estimated by measuring the distance the echo moves in a given time.

Airborne Interception Radar (AI)

When the fighter is nearing the enemy aircraft its own radar equipment is used for the final stage of the interception. The equipment designed for this purpose does not need a large maximum range (up to 20 miles may be sufficient) but, as we have seen, it must have a *very short minimum range*, i.e. the 'blind' area surrounding the radar must be as small as possible to enable the radar to be effective down to very short ranges (about 100 feet). Because of this the pulse duration used must be *very short*, $0.2 \mu s$ being typical. The short pulse duration also gives improved discrimination in range so that the fighter can tell whether there is more than one enemy aircraft on the same track.

The radar must also be capable of showing the relative bearing of the target to port or starboard of the fighter's heading and indicate its angle of elevation above or below the fighter.

The aerial assembly, or 'scanner', is a parabolic reflector mounted in the nose of the aircraft (the 'radome'). It produces a narrow pencil beam which can rapidly be moved for scanning (Fig. 8). While searching for the enemy aircraft the aerial beam has to scan a wide area ahead

AIRBORNE INTERCEPTION

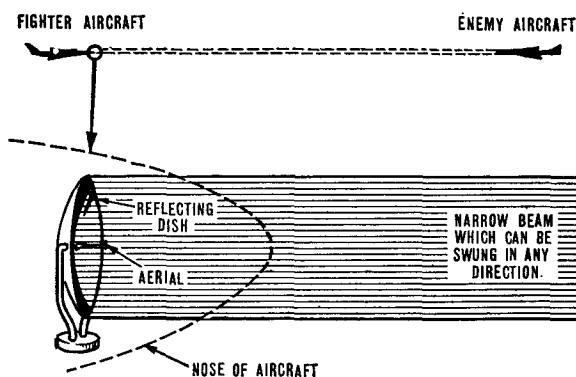


FIG 8. AI PENCIL BEAM

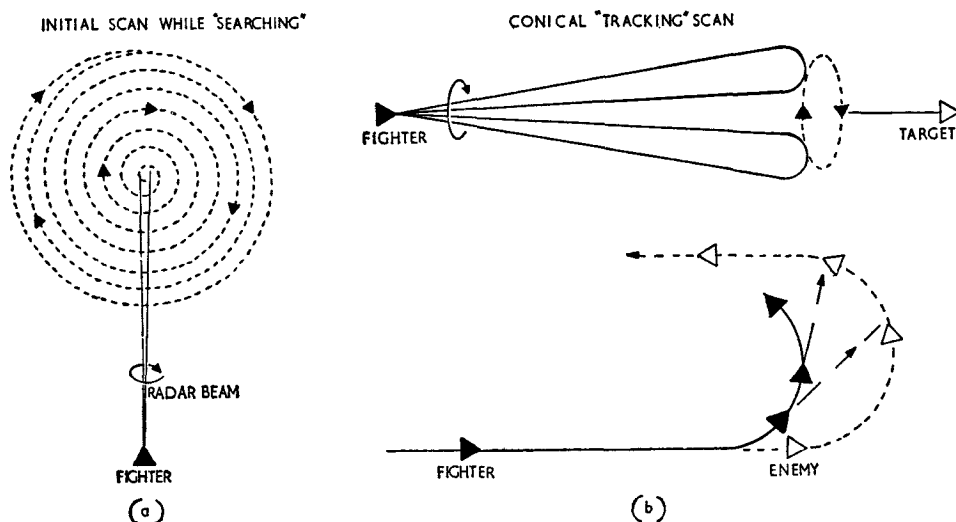


FIG 9. TYPICAL AI SCANNING METHODS

of the fighter; a *spiral scan* may be used for this (Fig 9a). When the selected target is picked up, the aerial may then be switched to *conical scan*.

The target echoes received during various portions of the conical scan are then used to operate circuits which give automatic 'following' or 'tracking' of the enemy aircraft (Fig 9b).

A typical AI indicator uses one c.r.t. in a type B display to show range and relative bearing (Fig 10). A bright line, called the range marker, is set by a control to cut the target echo and is then 'locked' so that as the range decreases and the bearing approaches dead ahead, the echo moves down the display and the range marker moves with it. In this condition a meter is automatically operated to give the range of the target in yards.

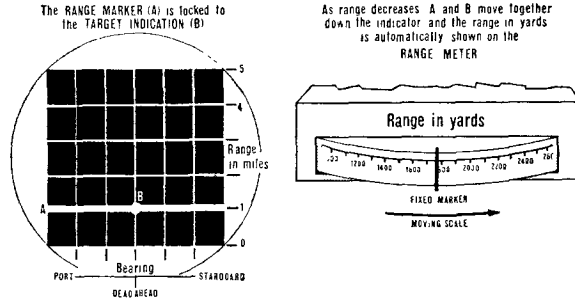


FIG 10. INDICATION OF RANGE AND RELATIVE BEARING

Another c.r.t. is used in a type F display to show bearing and elevation *errors* of the target (Fig 11). If the fighter pilot positions his aircraft so that the target indication is centred on the crossed lines in the type F display, the fighter is then pointing directly at the target. The range meter indicates when the enemy is within range of the fighter's armament and the hostile target can then be attacked even if the fighter pilot still cannot see it. It is in fact possible to arrange for

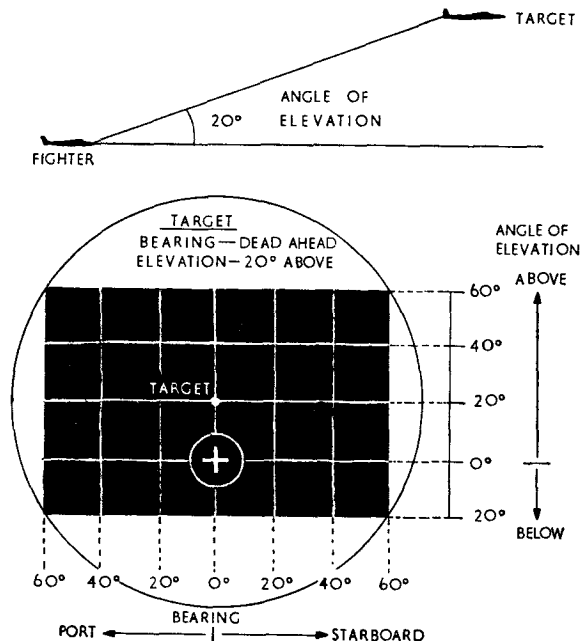


FIG 11. INDICATION OF BEARING AND ELEVATION ERRORS

the fighter's armament to be fired automatically. The radar circuits can be pre-set such that when the target echo is centred on the cross of the type F display and the range meter indicates, say, 800 yards the fighter's weapons are fired automatically.

Secondary Radar, Rebecca/Eureka

Let us assume that the fighter aircraft has destroyed the enemy and is returning to base. The weather may have 'closed in' to such an extent that the fighter pilot finds difficulty in locating his base. To enable aircraft to find their own airfield in conditions of poor visibility a secondary radar homing system may be used. An *interrogator* known as Rebecca is carried in the *aircraft*, and a *transponder* beacon known as Eureka is mounted in a fixed position on the *airfield*.

The aircraft uses a fixed transmitter aerial to radiate interrogating pulses ahead of the aircraft in a wide beam. Separate fixed receiver aerials pick up the transponder reply pulses from port or starboard. The radiation and reception patterns of these aerials are shown in Fig 12.

The pilot then has to steer only in the general direction of the airfield. When the aircraft is within radar range, transponder pulses are picked up by one or both of the receiver aerials. If the airfield is dead ahead, the amplitude of the pulses received by the port and starboard aerials will be equal; if the airfield is to starboard of the aircraft's heading the starboard aerial

receives the greater amplitude signal; if the airfield is to port the port aerial receives the greater amplitude signal (Fig. 13).

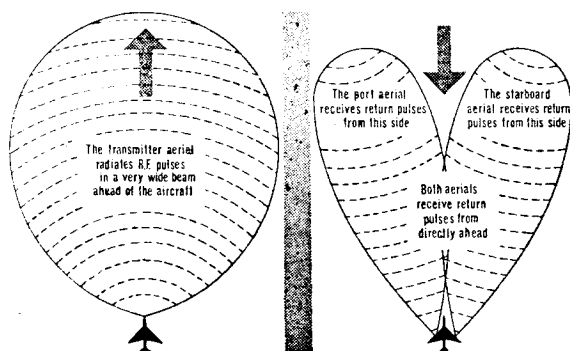


FIG 12. REBECCA RADIATION PATTERNS

When the EUREKA TRANSPONDER is in this position

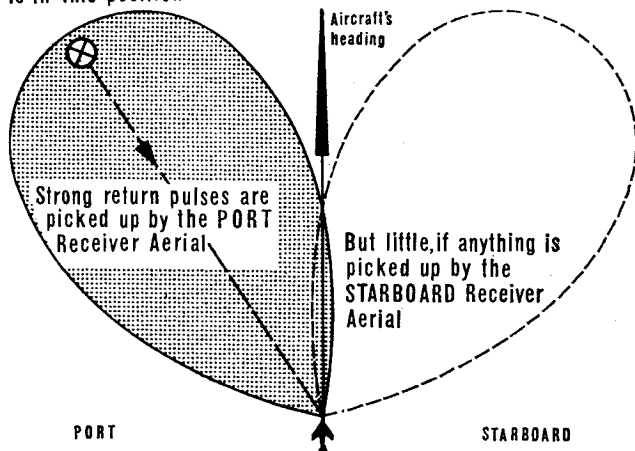


FIG 13. HOW BEARING IS OBTAINED

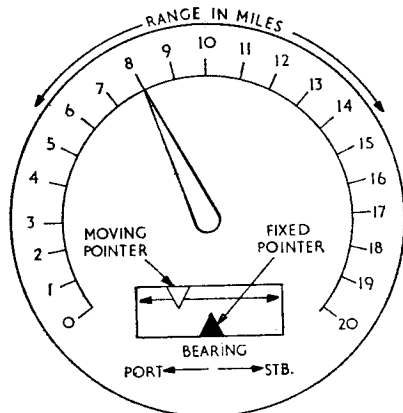
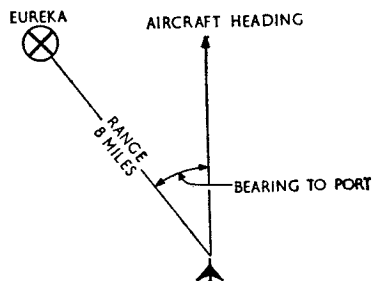


FIG 14. INDICATION OF RANGE AND BEARING

The received pulses could be displayed on a c.r.t. to show the bearing of the transponder beacon relative to the aircraft, but modern versions of Rebecca use a meter to convey the information directly to the pilot. Range is measured by timing the interval between transmission of the interrogating pulse from Rebecca and reception of the transponder pulse from Eureka. This can also be shown on a meter to give the pilot a simple instrument as shown in Fig 14. The reading shown in Fig 14 indicates that the airfield is slightly to port at a range of eight miles. The pilot then has to turn the aircraft to port until the two pointers of the bearing indicator are aligned. The aircraft is then heading directly for the airfield and the range is given on the outer scale.

Summary

Much more could be said about the various applications of pulse-modulated radar. However, the outlines given in the preceding paragraphs of this chapter indicate the various possibilities and some of these will be considered in greater detail in later chapters of this book. It is again emphasized that only the basic ideas, in very general terms, have been given in this chapter. Modern equipments are more complex than indicated here, and when working on such equipments the appropriate Air Publication should always be consulted.

CHAPTER 5

BASIC OUTLINE OF CW RADAR

Introduction

In an earlier chapter we noted that all pulse-modulated radars have a 'blind' area surrounding the installation within which no targets can be detected. The blind area depends upon the value of pulse duration but even with a pulse of $0.2 \mu\text{s}$ no target within 100 feet of the radar aerial can be detected. Sometimes it is necessary to detect and to measure the distance of targets from the radar aerial down to almost zero feet. Pulsed radar cannot be used for this; *frequency-modulated continuous wave radar* (f.m.c.w.) can.

The only way to measure the *speed* of a target in pulsed radar is to try to estimate the distance the echo on the c.r.t. screen moves in a given time. This is a rather indirect and inefficient method. A better method involves the use of unmodulated c.w. radar and the '*Doppler effect*'.

We are therefore concerned with two different forms of c.w. radar—f.m.c.w. and c.w. Doppler. In this chapter we shall consider the elementary ideas of both forms, illustrating the application of each with examples.

Frequency-modulated CW Radar

One application of f.m.c.w. radar is in aircraft altimeters. The normal *barometric* altimeter is operated by air pressure and has two limitations:

- If the atmospheric pressure changes while the aircraft is in flight the altimeter reading will change.
 - The barometric altimeter indicates height *above sea level*, or some other pre-set level. It does not tell the pilot his actual altitude *above the ground* (Fig 1).
- These limitations led to the development of the radar altimeter.

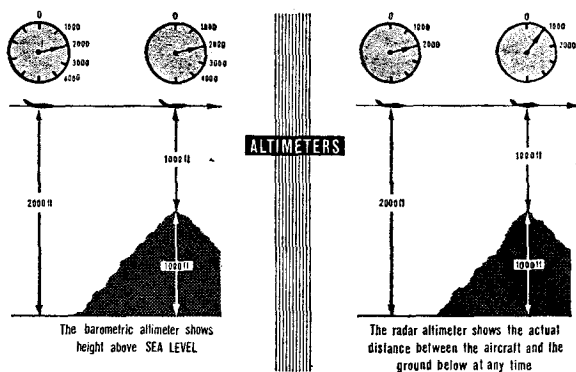


FIG 1. BAROMETRIC AND RADAR ALTIMETERS

We have seen how the distance between a radar aerial and a reflecting surface can be measured by pulse-modulated radar. If we transmit pulses directly downwards from an aircraft we can measure the *actual distance* to the ground below.

Altimeters which work on this principle give satisfactory results while the aircraft is at a high altitude. However, since all pulsed radars have a certain blind area, altimeters of this type would be useless when the aircraft is flying near the ground, e.g. when it is landing. For this we need a f.m.c.w. radar altimeter.

Radar Altimeters Using Frequency Modulation

The principles of frequency modulation are considered elsewhere in these notes (see pp 388 and 423 of AP 3302, Part 1B). Basically in a f.m. transmitter the carrier frequency is caused to change at a *rate* determined by the *frequency* of the modulating signal and by an *amount* determined by the *amplitude* of the modulating signal. The transmitter works *continuously* and produces a constant-amplitude c.w. output whose *frequency* is varied by the modulating signal.

Let us suppose that the frequency of a f.m. transmitter is caused to deviate at a *constant rate* by using a sawtooth waveform as the modulating signal (Fig 2). At point A the carrier frequency is, say, 400 Mc/s. At point B, 100 μ s later, the frequency is, say, 440 Mc/s. Since the change in frequency is *linear* we can say that the transmitter frequency is changing by 40 Mc/s every 100 μ s. Let us see how this principle is applied in the f.m.c.w. radar altimeter.

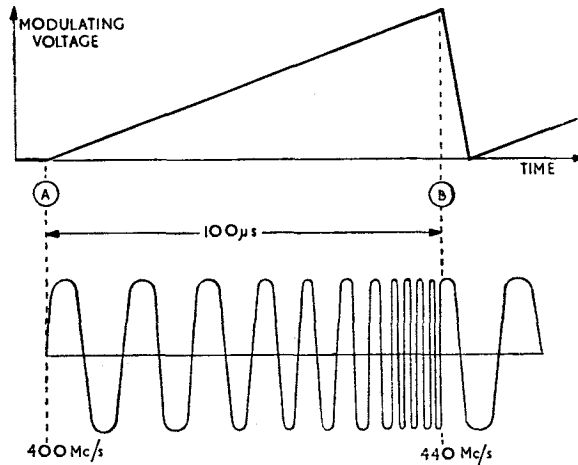


FIG 2. FREQUENCY MODULATION WITH A SAWTOOTH MODULATING SIGNAL

Fig 3 illustrates the layout of a typical f.m.c.w. radar altimeter in an aircraft. Let us assume that the output frequency is changing as described above and that at a given instant of time it is 410 Mc/s. The wave of this frequency is radiated downwards and reflected from the ground to be picked up by the receiver aerial. The wave takes a definite *time* to travel over this path so that

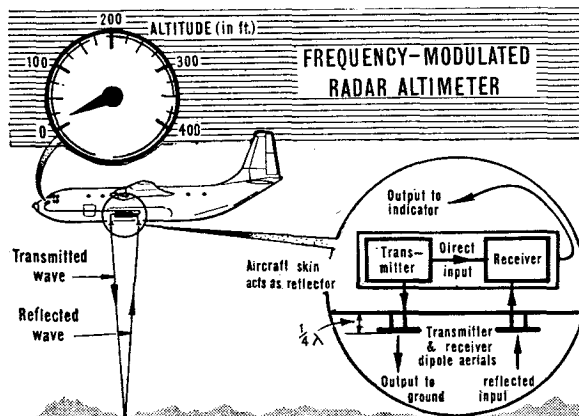


FIG 3. TYPICAL FMCW RADAR ALTIMETER LAYOUT

when it arrives back at the aircraft the *transmitter* frequency has in the meantime *changed* to, say, 410.2 Mc/s. The reflected wave of course has its *original* frequency of 410 Mc/s.

A portion of the transmitter output is fed *directly* to the receiver where it combines with the reflected input to produce a *difference* frequency, in this case 0.2 Mc/s. The *greater* the altitude of the aircraft the *greater* is the difference in frequency between the direct and reflected inputs. This difference frequency is automatically measured in discriminator circuits in the receiver, the output from which operates a simple meter display as shown in Fig. 3.

Since the transmitter frequency is changing *linearly* by 40 Mc/s every 100 μ s, a change of 0.2 Mc/s in the transmitter frequency represents a *time interval* of:

$$\frac{100 \times 0.2}{40} = 0.5 \mu\text{s}.$$

What range, or altitude, does a time interval of 0.5 μ s represent? We know that one radar mile (5,280 feet) is equivalent in time to 10.75 μ s. A time interval of 0.5 μ s therefore represents an altitude of:

$$\frac{5,280 \times 0.5}{10.75} = 250 \text{ feet approximately.}$$

This is merely one application of f.m.c.w. radar and it will be considered in more detail in Section 4.7 In general we can say that f.m.c.w. radar can be used to *detect* an object—indicated by the production of a difference frequency (a beat frequency) in the receiver discriminator circuits; it can measure the *range* of a target by measuring the beat frequency; and it can provide information on the *bearing* in azimuth and elevation of the target by using beamed radiation in the same way as pulsed radar.

The Doppler Effect

If we transmit a continuous wave at a fixed frequency, when the beam strikes an aircraft some of the r.f. energy is reflected (Fig 4). From the reflected signal, information about the

IF A CONTINUOUS WAVE IS TRANSMITTED FROM GROUND TO AIRCRAFT SOME OF THE ENERGY IS REFLECTED BACK TO THE GROUND

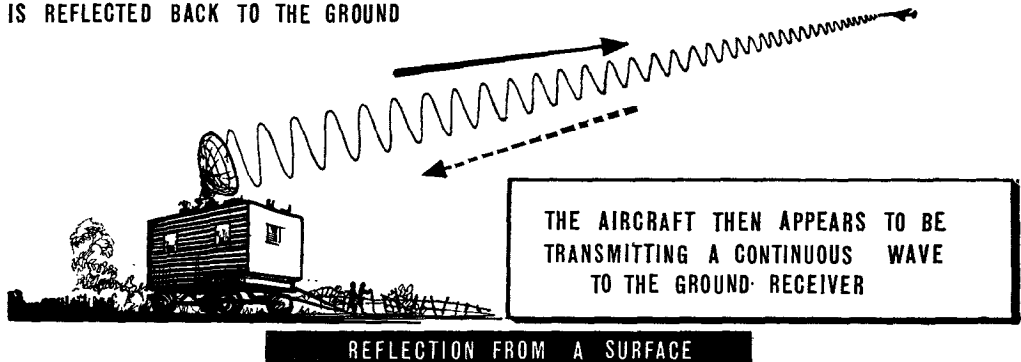


FIG 4. REFLECTION OF RADIO WAVES FROM A SURFACE

presence of a target and the target's angular position relative to the transmitting aerial can be obtained.

There is however another phenomenon associated with all wave propagation which is used in unmodulated c.w. radar.

If you are watching a motor-cycle race you may notice that the note of the exhaust noise appears to change as the machine passes. This is illustrated in Fig 5.

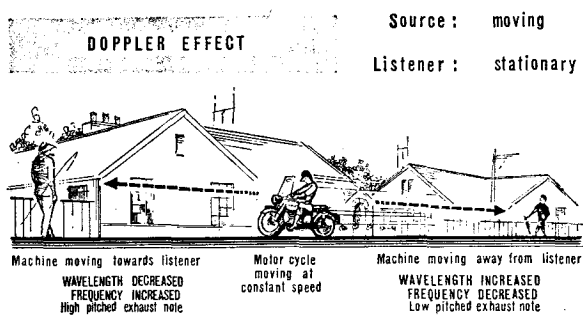


FIG 5. DOPPLER EFFECT WITH SOUND WAVES

This phenomenon is known as the *Doppler effect* and it occurs with radio waves as well as with sound waves. As a target *approaches* a radar aerial the *frequency* of the signal reflected by the target is *higher* than that of the transmitted signal. Conversely if a target is *moving directly away* from the aerial the frequency of the reflected signal is *lower* than that of the transmitted signal. For stationary targets there is *no change* in the frequency of the reflected signal (Fig 6).

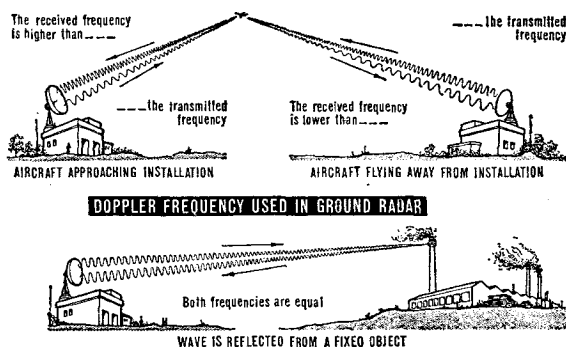


FIG 6. DOPPLER EFFECT WITH RADIO WAVES

If the transmitted frequency is f_t and the new frequency to which it is changed by the Doppler effect is f_r , the *difference* between these two frequencies is known as the *Doppler shift* $f_d = f_t - f_r$.

The *magnitude* of the Doppler shift is related to the *velocity* of a target in a straight line between the target and the aerial. A *high* value for the Doppler shift indicates a *high* target velocity.

If the target is *approaching* the aerial the received frequency is *higher* than the original transmitted frequency by the Doppler shift, i.e. $f_r = f_t + f_d$. If the target is *moving away* the received frequency is *lower*, i.e. $f_r = f_t - f_d$.

The relationship between a target's velocity and the Doppler shift, provided the target is approaching or receding in a straight line from the radar aerial, is given by the expression:

$$f_d \approx \frac{2v}{c} f_t,$$

- where f_d = Doppler shift in c/s
 f_t = Transmitted frequency in c/s
 v = Velocity of target in m.p.h.
 c = Velocity of radio waves in m.p.h.

If the transmitted frequency is 1,860 Mc/s and the velocity of a target directly approaching the aerial is 360 m.p.h. then:

$$\text{Doppler shift } f_d \approx \frac{2 \times 360}{186,000 \times 60 \times 60} \times 1,860 \times 10^6 = 2 \text{ kc/s.}$$

This means that the frequency of the received signal f_r is $f_t + f_d = 1,860 \text{ Mc/s} + 2 \text{ kc/s}$. If the target had been *moving away* in a direct line at 360 m.p.h. the frequency of the received signal would have been $f_r = f_t - f_d = 1,860 \text{ Mc/s} - 2 \text{ kc/s}$.

In practice it is the *velocity* of a target we wish to find, so we work the *other way round* from the measured value for the Doppler shift and the other known factors. Knowing the relationship it is simple to convert any *difference* in frequency between the received signal and the transmitted signal into the relative velocity of the target.

So far we have assumed that the target is moving *in a direct line* either towards or away from the radar aerial. If the target is not moving along such a path, the difference in frequency which Doppler effect causes is *less*. From Fig 7 we can see that the important factor is the *radial velocity*, i.e. that component of the target's speed which is *in a direct line* with the aerial. When

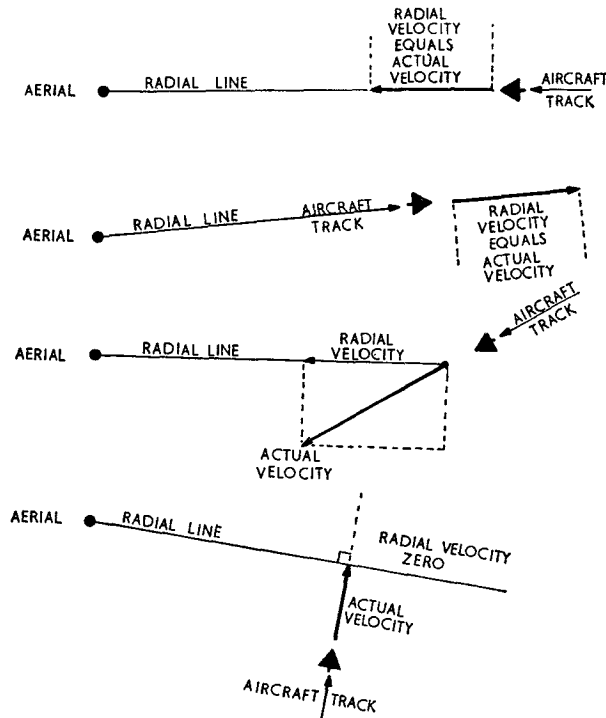


FIG 7. RADIAL VELOCITY

If the target is not moving along a radial line the radial velocity is *less* than the actual velocity. In fact if the target is moving *at right angles* across a radial line its radial velocity is *zero*. It is only the radial velocity which can be measured by the Doppler effect.

Use of the Doppler Effect in Ground Radar

With pulse-modulated ground radar equipment reflections from large fixed objects cause *permanent echoes* on the indicator, and random reflections from small objects close to the radar cause *clutter* at the centre of the p.p.i. (Fig 8).

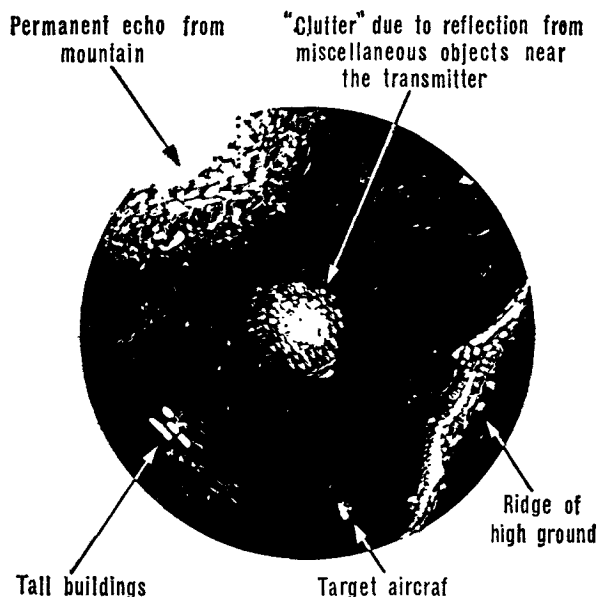


FIG 8. ECHOES FROM UNWANTED OBJECTS IN PULSED RADAR

For most applications the receiver should ignore reflections from *fixed* objects and respond only to *moving* targets. This can be achieved by using the Doppler effect because a Doppler frequency is produced only by the radial velocity of a *moving* target. For a stationary object the reflected signal has the same frequency as the transmitted signal.

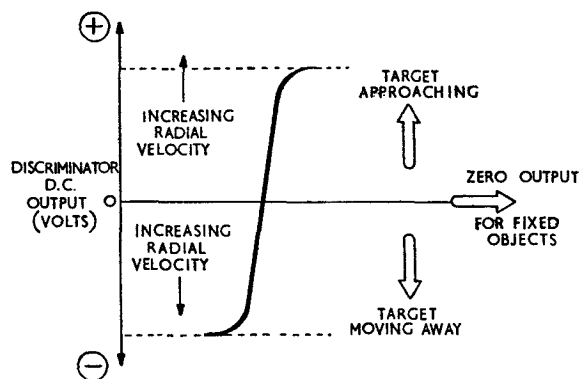
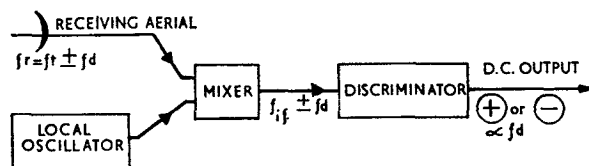


FIG 9. INDICATION OF MOVING TARGETS

Fig 9 illustrates a typical arrangement for the indication of a moving target. The frequency f_r of the reflected signal differs from that of the transmitted signal f_t by the Doppler shift f_d . The reflected signal is mixed with the output of a local oscillator to produce an i.f. signal $(f_{if} \pm f_d)$ where f_d is the Doppler shift. This signal is amplified and fed to a discriminator whose output is either a positive-going or a negative-going d.c. voltage depending upon whether the frequency of the reflected signal is *above* or *below* that of the transmitter. Remember that the frequency of the reflected signal increases if the target is approaching and decreases

if it is flying away from the radar. Thus the *sign* of the discriminator output indicates whether the target is approaching or moving away. The *magnitude* of the discriminator output depends upon the frequency deviation of the reflected signal in relation to the transmitter frequency, and this in turn is proportional to the *target's radial velocity*.

An installation using c.w. Doppler radar will provide the following information about a target:

- The *presence* of a *moving* target is indicated by the production of a Doppler frequency. Stationary objects provide no change in frequency.
- The *bearing* and *elevation* of the target is determined by using narrow beams.
- The *radial velocity* of the target is determined by measuring the Doppler shift in frequency.
- The *direction of travel* of the target is determined by noting the sign of the Doppler shift.

Note that c.w. Doppler *does not measure the range* of a target. For this we use either pulse-modulated radar or f.m.c.w. radar.

Use of the Doppler Effect in Airborne Radar

If the radar transmitter is located in an aircraft the signals reflected from the ground ahead of the aircraft will also be subject to the Doppler effect. Use is made of this property in aircraft navigation. To navigate accurately one important factor which must be known by the navigator is the *ground speed* of the aircraft. This may be quite different from the *air speed*. Let us see how the Doppler effect helps here.

Since the ground ahead of the aircraft is being illuminated by the radar beam from the airborne transmitter it will reflect energy back towards the aircraft. The aircraft is always *moving towards* the apparent source of radiation and so the received frequency f_r is *higher* than the transmitted frequency f_t by the Doppler shift f_d , i.e. $f_r = f_t + f_d$. The Doppler shift is determined by the radial velocity of the aircraft and is given as before by the expression $f_d \approx \frac{2v}{c} f_t$.

Special circuits in the receiver automatically measure the Doppler shift and the receiver output can be displayed on a simple meter calibrated in m.p.h. A practical equipment which uses this system transmits not one radar beam but *four* at different points around the aircraft. The information which is received from all four points on the ground is used to eliminate errors which would otherwise arise when the aircraft is climbing, diving or banking. The four beams also enable the *drift* of the aircraft to be calculated. A typical meter display is illustrated in Fig 10.

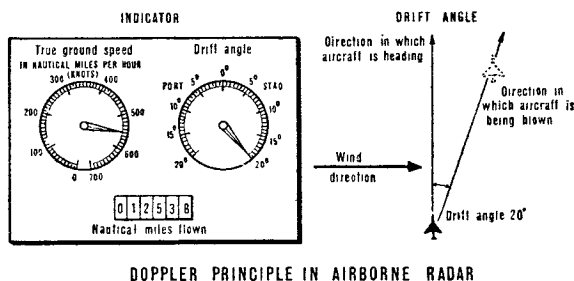


FIG 10. DOPPLER AIRBORNE NAVIGATION DISPLAY SYSTEM

Summary

In this chapter a broad outline of the ideas involved in the use of c.w. radar has been given. Many factors have not yet been considered, nor can they be until we have learned something

about the operation of radar circuits and components. Further details of c.w. radar will therefore be found in Section 7.

It should be noted that the frequency of received pulses in a *pulsed radar* is also altered by Doppler shift. However, the Doppler shift is small and in a pulsed radar is accepted within the bandwidth of the receiver circuits. In a Doppler radar the shift frequency is the important factor and to ensure accuracy the frequency stability of a Doppler radar must be high.

This concludes the first section of this book. The aim was to show the basic principles of radar and its limitations and applications. Having learned how basic systems operate at block level we must now proceed to 'look inside' the various boxes to see what they contain and how they function to provide the outputs we need. This is done in succeeding sections.

SECTION 2

RADAR CIRCUITS

Chapter	Page
1 Square Waves	61
2 Square Waves Applied to CR Circuits	67
3 Square Waves Applied to LR and LC Circuits	77
4 Limiting Circuits	83
5 Clamping Circuits	95
6 Free-running (Astable) Multivibrators	107
7 Monostable and Bistable Multivibrators	123
8 Other Square Wave Generators	135
9 Ringing and Blocking Oscillators	143
10 Electronic Switching Circuits	155
11 Frequency-dividing and Counting Circuits	165
12 Timebase Principles	175
13 Miller Timebase Circuits	191
14 Other Radar Timebase Generators	201
15 Strobe Pulse Circuits	209
16 Paraphase Amplifiers	219

CHAPTER 1

SQUARE WAVES

Introduction

In Section 1 we learned the basic principles of operation of a pulse-modulated radar system and saw how the system may be built up in the form of a block schematic diagram where each block represents a group of special circuits. Before we can tackle the job of maintaining such equipment we must know more than the *purpose* of each block; it is important to understand *how it produces* the required results. We must therefore 'look inside' each block and learn about the special circuits which work together to make up a complete radar installation (Fig 1). In this section we shall look at some of these basic radar circuits.

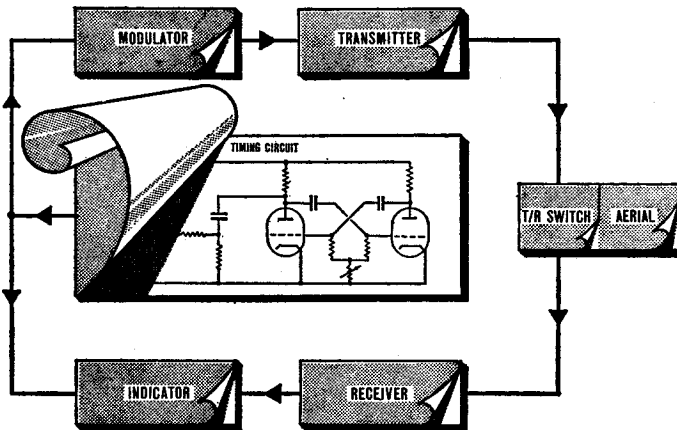


FIG 1. FROM BLOCKS TO CIRCUITS

We know that a radar transmitter must be switched on and off many times each second to produce a series of pulses of r.f. energy. We also know that the indicator timebase circuits must be switched on at the instant each transmitter pulse begins and switched off after a time equal to the maximum required radar range. As we progress we shall find that many other circuits must be switched on at the *instant* the transmitter fires or an *exact* number of microseconds later.

Mechanical switches and relays cannot operate with the precision which is essential for this type of circuit switching. For accurate synchronization within the equipment the switching must be done *electronically* with *square waves* of voltage. We shall see later that almost all pulsed radar circuits are concerned with the production or transmission of such waveforms. Let us therefore begin by learning what is meant by the term 'square wave'.

The Simple Square Wave

A waveform is a graph which shows how a voltage or a current varies over a given period of time. A square wave is a voltage or a current change in which the waveform has *square*, i.e. right-angled, corners.

Fig 2a shows a simple circuit of a battery, a switch and a bulb, in which the negative terminal of the cell is earthed. The voltage at the positive terminal is therefore $+1.5\text{V}$ with respect to earth. If now the circuit is switched alternately on and off at intervals of two seconds, the waveform of voltage applied to the bulb is a *square wave* as shown in Fig 2b.

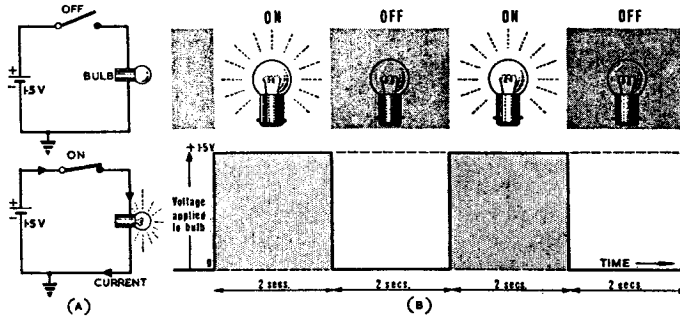


FIG 2. ELEMENTARY SQUARE WAVE 'GENERATOR'

Amplitude and Polarity of Square Waves

The square wave of Fig 2b starts from zero and rises to $+1.5\text{V}$. Its *amplitude* is therefore 1.5V . To describe its *polarity* we say that it is *positive-going* with respect to earth.

Somewhat similar results would be obtained if the battery were reversed so that its *positive* terminal was earthed, but in this case the voltage applied to the bulb would be -1.5V with respect to earth. The square wave of voltage produced by switching on and off would then be as shown in Fig 3. The voltage starts from zero and falls to -1.5V . The amplitude of the waveform is again 1.5V but its polarity is now *negative-going* with respect to earth.

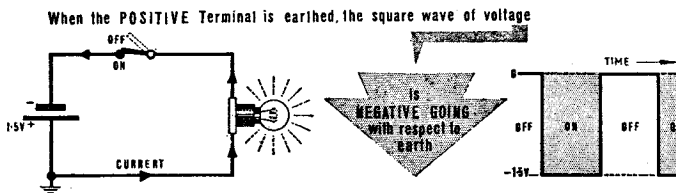


FIG 3. PRODUCTION OF NEGATIVE-GOING SQUARE WAVE

The square waves we use in radar do not always start from zero but may vary between any two fixed voltage levels. Some examples are shown in Fig 4. In each case the *amplitude* of the

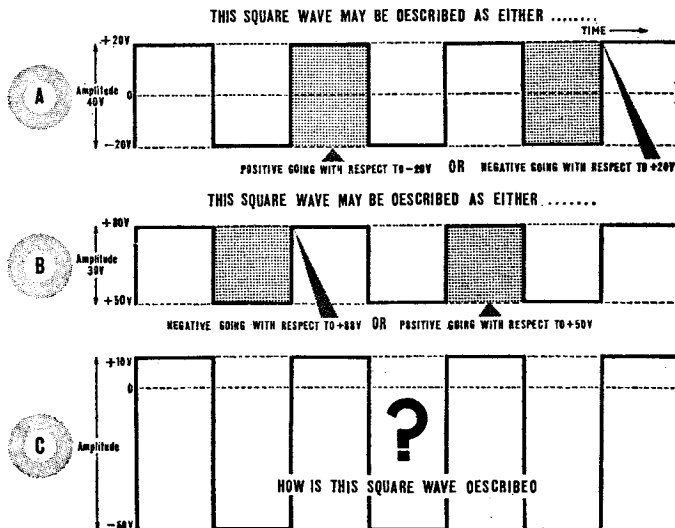


FIG 4. HOW SQUARE WAVES ARE DESCRIBED

square wave is the *difference* between the two voltage levels. The *polarity* may be stated as either positive-going or negative-going with respect to one of the fixed voltage levels. In most cases the reference level will suggest itself naturally, e.g. we often use h.t. + as a reference level.

Period and Frequency of Square Waves

The *period* of the square waves used in radar equipment is usually expressed in microseconds. For example the period of the square wave shown in Fig 5 is 200 microseconds. The *frequency* of this waveform is therefore 5,000 cycles per second.

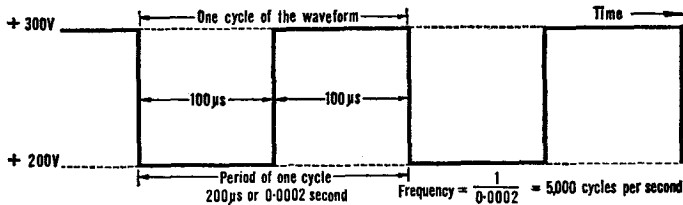


FIG 5. PERIOD AND FREQUENCY OF A SQUARE WAVE

Symmetry

So far we have considered only the case where the time interval during which the circuit is switched on is *equal* to that during which it is switched off. The two parts of the waveform are then of *equal duration* and the square wave is said to be *symmetrical* (see Fig 6a).

If the time interval during which the switch is open is longer or shorter than the time during which it is closed the two parts of the waveform are *no longer of equal duration*. The resultant waveform is often called an *asymmetrical square wave* or a *rectangular wave* (Fig 6b and c).

These various terms are often used indiscriminately and any of the waveforms shown in Fig 6 may be loosely referred to as either square waves or rectangular waves.

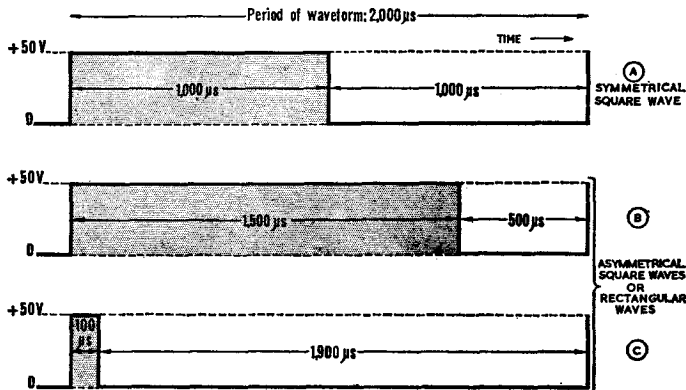


FIG 6. SYMMETRICAL AND ASYMMETRICAL SQUARE WAVES

Pulse Duration and PRF

The rectangular wave shown in Fig 6c is the waveform most commonly met with in pulsed radar. To indicate the length of each part of one cycle we call either part a *pulse* and give the *pulse duration* in microseconds.

A pulse may be defined by the terms 'narrow' and 'wide' (see Fig 7). A *narrow pulse* is one in which the pulse duration is very short compared with the period of the rectangular wave.

The pulse duration of a *wide* pulse, on the other hand, is a large fraction of the period of the waveform.

The *leading* and *trailing* edges of a pulse are also illustrated in Fig 7. If the leading edge *rises* from the reference level the pulse is *positive-going* (Fig 7a). If the leading edge *falls* from the reference level the pulse is *negative-going* (Fig 7b).

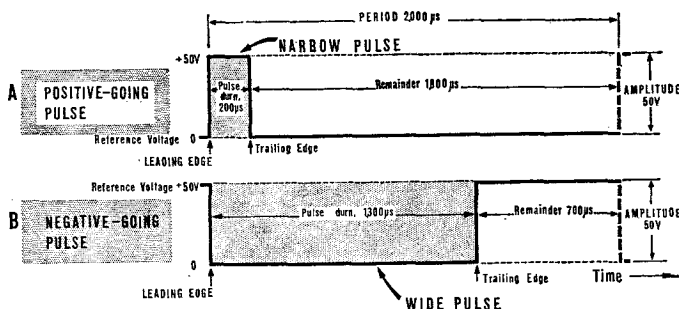


FIG 7. PULSE CHARACTERISTICS

We have seen in Section 1 that the *pulse repetition frequency* (p.r.f.) is the *number* of pulses occurring in one second. In both parts of Fig 7 we have one pulse every 2,000 μ s. Therefore in one second we have $\frac{10^6}{2,000}$ or 500 pulses. This is the p.r.f.

Pulse Shape and Bandwidth

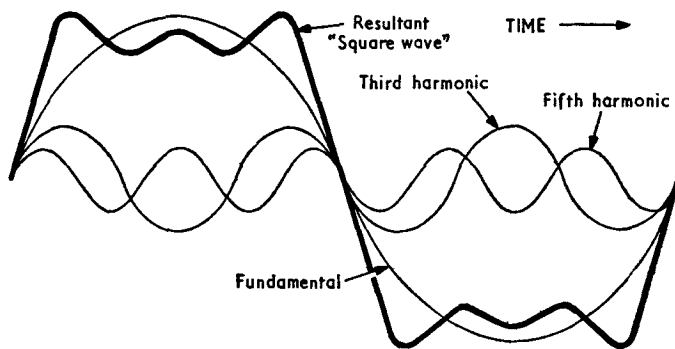
The square waves considered so far have all been *ideal* waveforms, i.e. with zero rise and fall times and with perfectly flat tops. Such a square wave is made up of *sine waves* of a fundamental frequency equal to that of the square wave plus an *infinite* number of *odd* harmonics, all starting in phase. In practice, of course, no circuit is capable of passing an *infinite* number of harmonics so that the ideal waveform described cannot exist. However the *greater* the number of odd harmonics contained in the waveform the more closely it approaches the ideal waveform and the *steeper* are the leading and trailing edges (see Fig 8).

For accurate timing and synchronization within a radar equipment, and for accurate range measurement, pulses with *steep* leading edges are essential. This means that the pulse must contain a *large number* of harmonics. The rise time of the leading edge of the pulse must be *short* compared with the pulse duration—not greater than about *one-tenth* of the pulse duration. For a pulse whose duration is 1 μ s the rise time should not exceed 0.1 μ s.

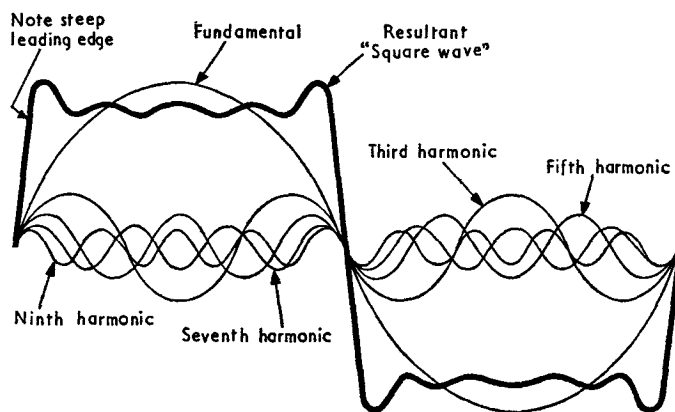
To a good approximation this rise time t can be taken as the time of one quarter cycle of the *highest harmonic* within the pulse. Thus the time of one cycle of the highest harmonic is $4t$, giving a frequency of $\frac{1}{4t}$ cycles per second (Fig 9). If the rise time is 0.1 μ s the frequency of the highest harmonic within the pulse is:—

$$\frac{1}{4t} = \frac{1}{4 \times 0.1 \times 10^{-6}} = 2.5 \text{ Mc/s.}$$

When a carrier of frequency f_c is amplitude-modulated by a sine wave of frequency f_m , sidebands $f_c + f_m$ and $f_c - f_m$ are produced. A radar transmitter is a sine wave oscillator modulated by a *pulse* waveform and hence the radiated r.f. pulse consists of the carrier frequency $\pm$ all the harmonic components of the modulating pulse. In the above example the frequency of the highest harmonic is 2.5 Mc/s so that the *band* of frequencies in the radiated r.f. pulse extends from $f_c - 2.5 \text{ Mc/s}$ to $f_c + 2.5 \text{ Mc/s}$.



(A) USING UP TO FIFTH HARMONIC



(B) USING UP TO NINTH HARMONIC

FIG 8. CONSTRUCTION OF SQUARE WAVES FROM SINE WAVES

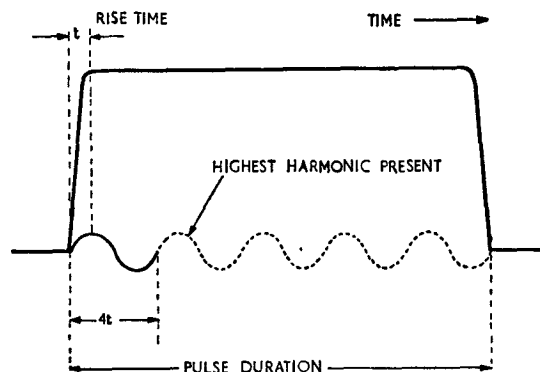


FIG 9. RISE TIME OF PULSE AND BANDWIDTH

To preserve the *shape* of the pulse at the receiver the *bandwidth* of the r.f. and i.f. amplifiers must be sufficient to amplify all frequencies extending 2.5 Mc/s *either side* of the centre frequency, i.e. in this example the r.f. and i.f. amplifiers must have a *bandwidth* of 5 Mc/s to avoid distortion.

To maintain the same *percentage* distortion for a pulse of *shorter duration* the leading edge of the pulse must rise in a *much shorter* period of time. For example if the pulse duration is reduced to $0.5 \mu\text{s}$ the leading edge of the pulse must rise in a time t not greater than $0.05 \mu\text{s}$. The frequency of the highest harmonic within the pulse is now:—

$$\frac{1}{4t} = \frac{1}{4 \times 0.05 \times 10^{-6}} = 5 \text{ Mc/s.}$$

The r.f. and i.f. bandwidths in the receiver must now be increased to 10 Mc/s. Thus as the pulse duration is *reduced* and the leading edge of the pulse made *steeper* the *bandwidth* of the circuits handling the pulse must *increase* accordingly. We shall see in a later chapter how wideband amplifiers are obtained.

Summary

- a. A *square wave* of voltage starts from a fixed voltage level, changes to a different voltage in a negligibly short time, remains at this new voltage for a certain time and then changes to the original voltage level in a negligibly short time, remaining there for some time.
- b. The *amplitude* of a square wave is the difference between the two voltage levels between which the waveform varies.
- c. The *polarity* is described by taking one of the voltage levels as a reference and stating whether the wave goes positive or negative with respect to that level.
- d. The *period* of a square wave is the time taken by one complete cycle of variation.
- e. The *frequency* of a square wave is the number of cycles occurring in one second.
- f. A *symmetrical* square wave is one in which the two parts of the cycle are of *equal* duration.
- g. An *asymmetrical* square wave (a *rectangular wave*) is one in which the two parts of the cycle are of *unequal* duration.
- h. Either part of one cycle of a square wave can be called a *pulse* and its duration is called the *pulse duration* (sometimes pulse length or pulse width). The terms 'wide' and 'narrow' are also used to describe a pulse.
- j. The *number* of pulses occurring each second defines the *pulse repetition frequency*.
- k. For accurate timing and range measurement, pulses with *steep leading edges* are necessary. Because of the large number of harmonics associated with a steeply rising pulse *wideband* circuits are then needed.

CHAPTER 2

SQUARE WAVES APPLIED TO CR CIRCUITS

Introduction

In Chapter 1 we showed that square and rectangular waves may be used to switch radar circuits on and off at precise instants of time. However, square waves are used in radar for many other purposes also. Many radar circuits actually depend for their operation upon the effects of applying a square wave to a capacitor-resistor (CR) circuit.

We already know that if a sine wave input is applied to a CR circuit the output waveform is also a sine wave (Fig 1). However if we apply a square wave input to a CR circuit the output is *not necessarily* a square wave. In this chapter we shall examine the effects of applying square waves to CR circuits.

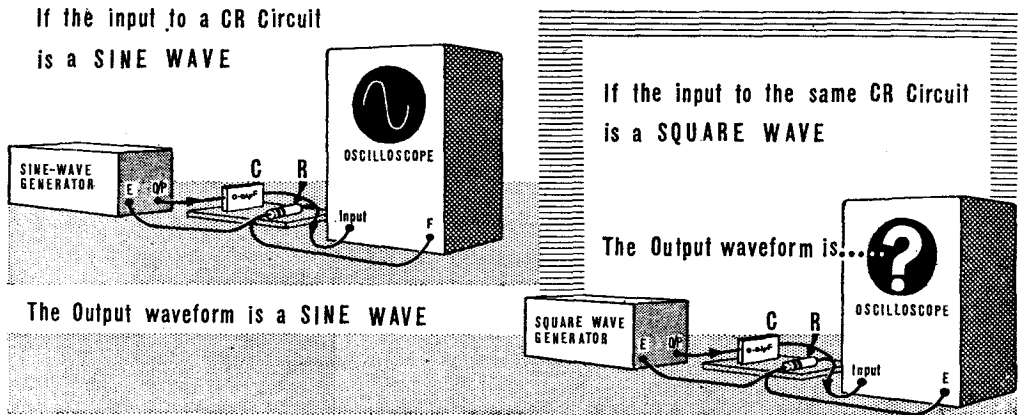


FIG 1. SINE WAVE AND SQUARE WAVE APPLIED TO A CR CIRCUIT

CR Time Constant

We already know from Part 1 of these notes that when a resistor R is connected in series with a capacitor C across a source of voltage:

- The charge on the capacitor cannot change *instantaneously* so that initially the full source voltage is developed across R .
- The capacitor then commences to charge *exponentially* at a rate dependent upon the circuit time constant, CR seconds. As the voltage V_C across C rises so the voltage V_R across R falls.
- After a time equal to CR seconds the capacitor is two-thirds (63 per cent) charged and is assumed to be fully charged after $5 CR$ seconds.
- At all times the sum of the voltages across C and R equals the applied voltage (Kirchhoff's second law), i.e. $V_C + V_R = V$.

Similarly when C is allowed to discharge through R , V_C falls exponentially from the charged voltage to zero in a time of $5 CR$ seconds. Thus when we apply a *square wave* of voltage to a CR circuit, V_C will rise and fall exponentially and V_R will vary correspondingly to maintain the relationship $V_C + V_R = V$. Let us now consider this action in more detail.

Effect of Applying a Square Wave to a CR Circuit

In Fig 2 we have a CR circuit and a simple 'square wave generator', i.e. the battery and the switch. Let us start from a point when the switch is in the *discharge* position and C is fully

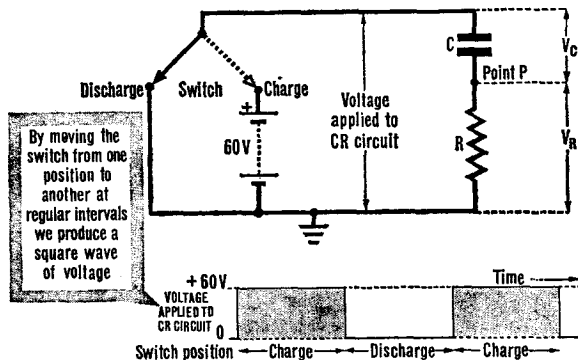
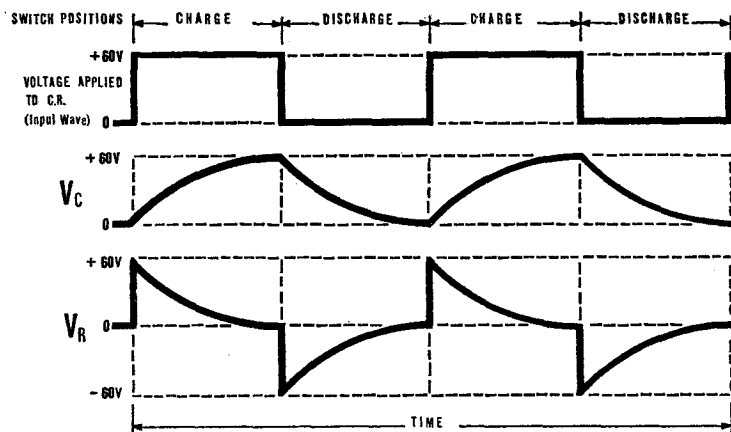


FIG 2. CIRCUIT FOR APPLYING A SQUARE WAVE TO A SERIES CR CIRCUIT

discharged. The applied voltage V is then zero, V_C is zero and, since there is no current flowing through R , V_R is also zero.

When we switch to the *charge* position, C cannot change its charge—and hence its voltage— instantaneously, so that the *whole* of the applied voltage V appears across R . Thus V_R rises *immediately* to the voltage V of the supply, in this case $+60V$; V_C at this instant is zero (Fig 3).

FIG 3. V_C AND V_R IN A CR CIRCUIT, SQUARE WAVE INPUT

Subsequently V_C begins to rise and will continue to rise exponentially until C is fully charged to $+60V$ in a time of $5CR$ seconds. Since at all times $V_C + V_R = V$, V_R falls exponentially towards zero as V_C rises.

At the instant when C is fully charged to the voltage of the supply ($+60V$), and V_R is consequently zero, suppose we switch back to the *discharge* position. This causes the applied voltage V to fall immediately from $+60V$ to zero. Remembering that C cannot change its charge instantaneously V_C will initially be unaffected by this and remain charged to $+60V$. But what of V_R ? Since $V = V_C + V_R$, and the applied voltage V is zero, we have at this instant:

$$0 = V_C + V_R$$

$$\therefore V_R = -V_C = -60V.$$

Thus at the instant of switching to the discharge position the voltage V_R across R falls *immediately* to $-60V$. (The negative sign means that point P in the circuit of Fig 2 is *negative* with respect to earth).

Subsequently C discharges through R causing V_C to *fall* exponentially from + 60V towards zero and V_R to *rise* exponentially from - 60V towards zero (Fig 3).

As we can see from Fig 3 the period of the square wave applied to the CR circuit is such that C can *just charge or discharge* completely during each half-cycle. We can also see that the waveforms of voltage across C and across R are *not square waves*.

Variations in the Basic Circuit

If a sudden voltage (a voltage 'step') is applied to a CR circuit in which C *already has an initial voltage*, the time constant is still CR seconds to the voltage step. In time CR seconds the voltage across C changes by 63 per cent of the *difference* between its initial voltage and the voltage step. Thus if C is charged initially to 20V, and 100V is then suddenly applied to the circuit,

V_C rises to $20 + \left(\frac{63}{100} \times 80\right) = 70.5V$ in CR seconds and to 100V in 5 CR seconds (Fig 4).

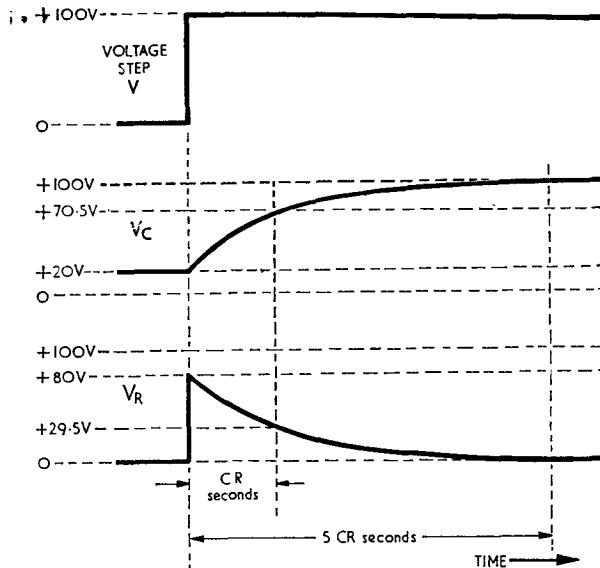


FIG 4. VOLTAGE STEP APPLIED TO A CR CIRCUIT, C INITIALLY CHARGED

As before, $V = V_C + V_R$. Hence at the instant the voltage step V is applied, $V_R = 80V$. It then falls to zero as V_C rises to 100V.

A circuit with two or more resistors connected in series with a capacitor behaves as if the resistors were replaced by a single resistor equal in value to their sum (Fig 5).

The formula $V = V_C + V_R$ now becomes:—

$$V = V_C + V_{R_1} + V_{R_2}.$$

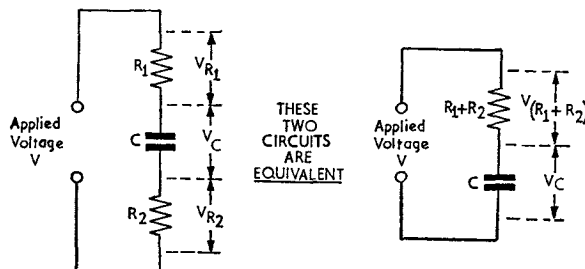


FIG 5. CR CIRCUIT WITH TWO RESISTORS

Since the same current is flowing through each resistor at any given instant, the voltages across the resistors will be in proportion to their resistances. For example, if 100V is applied to a CR circuit containing two series resistors valued at $10\text{k}\Omega$ and $90\text{k}\Omega$ then at the instant of applying the voltage the $10\text{k}\Omega$ resistor would have 10V across it and the $90\text{k}\Omega$ resistor 90V.

Suppose we have a circuit consisting of a $90\text{k}\Omega$ resistor and a $10\text{k}\Omega$ resistor connected in series with a $2\mu\text{F}$ capacitor which is initially charged to 20V (Fig 6a). The time constant CR is $2 \times 10^{-6} \times 100 \times 10^3 = 0.2$ second. If we now apply a voltage step of +100V to this circuit, V_C remains initially at +20V, because C cannot change its charge immediately. Since $V = V_C + V_{R_1} + V_{R_2}$, we then have 80V divided between R_1 and R_2 in proportion to their resistances. Across R_1 we have $\frac{90}{100} \times 80 = 72\text{V}$ and across R_2 we have $\frac{10}{100} \times 80 = 8\text{V}$ (Fig 6b).

C then commences to charge, rising to $20 + (\frac{63}{100} \times 80) = 70.5\text{V}$ in CR seconds (0.2 second) and to 100V in 5 CR seconds (1 second). Both V_{R_1} and V_{R_2} fall to zero as V_C rises to +100V. At all times the relationship $V = V_C + V_{R_1} + V_{R_2}$ is maintained.

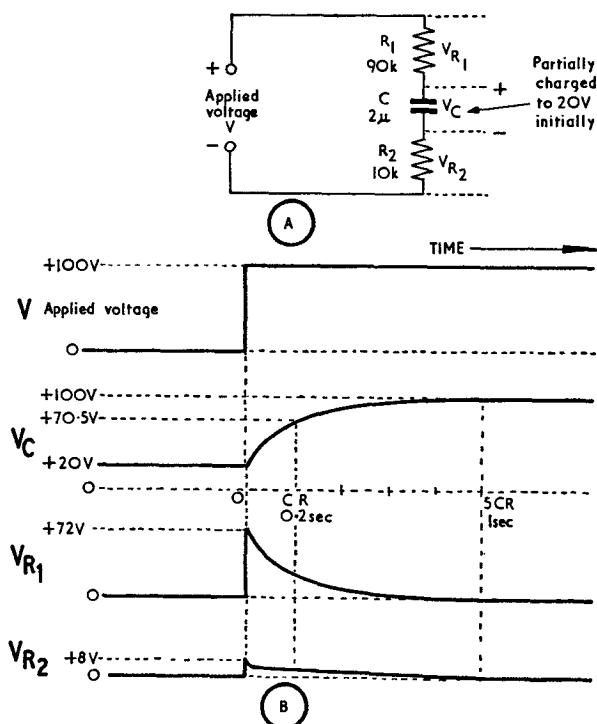


FIG 6. VOLTAGE STEP TO CR CIRCUIT, TWO RESISTORS, C INITIALLY CHARGED

Comparison of CR Time Constant with Pulse Duration

When the voltage input to a CR circuit is a square wave the exact shapes of the waveforms across C and across R depend upon how the CR time constant compares with the pulse duration of each part of one cycle of the square wave.

We have already seen the waveforms of V_C and V_R when the relationship between the CR time constant and the pulse duration is such that the capacitor can just charge and discharge in the time available (Fig 3). If the time constant is very short compared with the pulse duration, C can charge and discharge in a fraction of the available time and the waveforms of V_C and V_R are

then different from those in Fig 3. If the time constant is *very long* compared with the pulse duration, C can only *partially* charge and discharge in the time available and so we have another different set of waveforms for V_C and V_R .

In practice we compare the time constant of each CR combination with the pulse duration of the applied square wave as follows:

- If the time constant is *one-tenth* of the pulse duration (or *less*) we call the combination a *short CR* circuit.
- If the time constant is *ten times* the pulse duration (or *more*) we call the combination a *long CR* circuit.
- If the time constant is *between* these two extremes, i.e. between one-tenth and ten times the pulse duration, the combination is called a *medium CR* circuit.

Note that a circuit can be stated to be a short, medium or long CR circuit only in its relationship to a given input. A circuit which acts as a short CR to one input may well be a long CR to another.

Short CR Circuits

Let us assume that a symmetrical square wave of voltage, of amplitude 100V and pulse duration 1,000 μs , is applied to a CR circuit where $C = 0.01 \mu\text{F}$ and $R = 2,000 \Omega$ (Fig 7a).

Time constant $CR = 0.01 \times 10^{-6} \times 2,000 = 20 \mu\text{s}$.

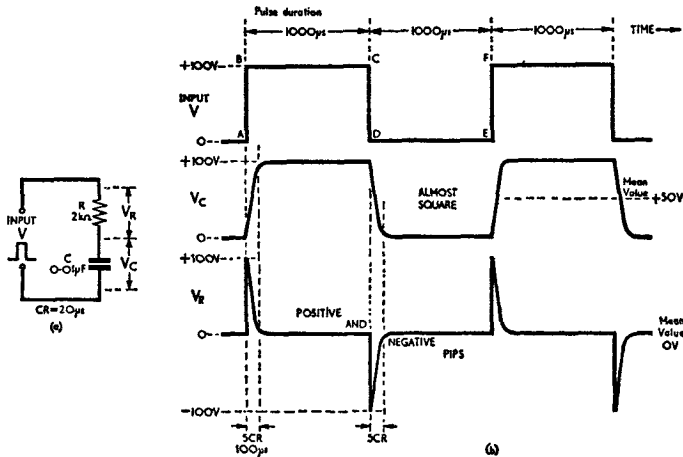


FIG 7. SQUARE WAVE APPLIED TO A SHORT CR CIRCUIT

This is *less than one-tenth* of the pulse duration so that the combination is a *short CR* to this particular input.

The resultant waveforms of V_C and V_R are shown in Fig 7b and are derived as follows:—

- A to B.* The input rises from zero to + 100V. Because C cannot change its charge instantaneously V_C remains at zero volts and V_R rises *instantly* to + 100V.
- B to C.* C commences to charge and V_C rises to + 100V in a time of $5CR$ (100 μs). V_R falls to zero in the same time, maintaining the relationship $V = V_C + V_R$.
- C to D.* The input falls by 100V from + 100V to zero and, since C is fully charged to + 100V and cannot change its charge instantaneously, V_R must also fall by 100V from zero to - 100V.
- D to E.* C discharges exponentially and V_C falls to zero in a time of $5CR$ (100 μs). V_R rises from - 100V to zero in the same time.
- E to F.* The cycle then repeats as from *a*.

Differentiating Circuit

There are many occasions in radar when we need to convert a square wave input to a series of positive- and negative-going pips of voltage. These pips may be used for timing purposes. They may be obtained by applying a square wave to a *short CR* circuit and taking the output across the *resistor*. A CR circuit used for this purpose is called a *differentiating circuit* (Fig 8). If the pulse duration of the input square wave is $200\ \mu\text{s}$ the CR time constant must be $20\ \mu\text{s}$ or less. Thus if the capacitor C is 100pF the resistor R must be $200\text{k}\Omega$ or less.

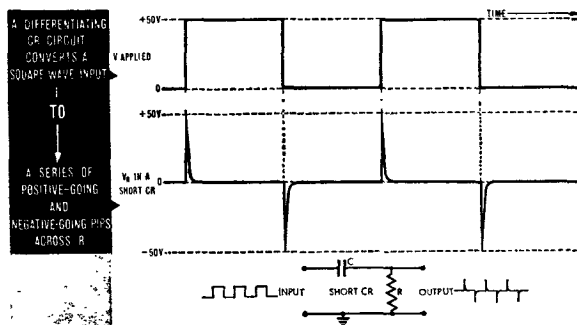


FIG 8. DIFFERENTIATING CIRCUIT

Note from Fig 8 that the positive-going pips are coincident in time with the rising edges of the square wave input, and the negative-going pips with the falling edges. In other words the pips are *linked in time* to the p.r.f. of the square wave input and this time linkage is important. We shall see later that either the positive-going pips or the negative-going pips may be eliminated by a limiter circuit, leaving us with pips all in one direction which may be used for triggering other circuits at precise instants of time.

Medium CR Circuit

Medium CR circuits are those in which the time constant is comparable with the pulse duration of each part of one cycle of the square wave input. Let us assume that a symmetrical square wave of voltage, of amplitude 100V and pulse duration $100\ \mu\text{s}$, is applied to a CR circuit where $C = 100\text{pF}$ and $R = 1\text{M}\Omega$ (Fig 9a).

Time constant $CR = 100 \times 10^{-12} \times 10^6 = 100\ \mu\text{s}$.

This is *equal* to the pulse duration so that the combination is a medium CR to this particular input.

The resultant waveforms of V_C and V_R are shown in Fig 9b and are derived as follows:—

a. *A to B.* The input rises from zero to $+100\text{V}$. Because C cannot change its charge instantaneously, V_C remains at zero volts and V_R rises *instantly* to $+100\text{V}$.

b. *B to C.* C commences to charge and V_C rises exponentially, reaching 63 per cent of 100V , i.e. $+63\text{V}$, in a time of CR ($100\ \mu\text{s}$). V_R falls to $+37\text{V}$ in the same time, maintaining the relationship $V_C + V_R = V = 100\text{V}$.

c. *C to D.* The input falls by 100V to zero and, since V_C is $+63\text{V}$ and cannot change immediately, V_R also falls by 100V from $+37\text{V}$ to -63V .

d. *D to E.* C discharges exponentially and in time CR ($100\ \mu\text{s}$) V_C falls by 63 per cent of 63V , i.e. V_C falls from $+63\text{V}$ to $63 - \left(\frac{63}{100} \times 63\right) = +23.5\text{V}$. V_R rises from -63V to -23.5V in the same time ($V_C + V_R = 0$ at this instant).

e. *E to F.* The input rises by 100V from zero so that V_R rises immediately by 100V from -23.5V to $+76.5\text{V}$. V_C remains at $+23.5\text{V}$.

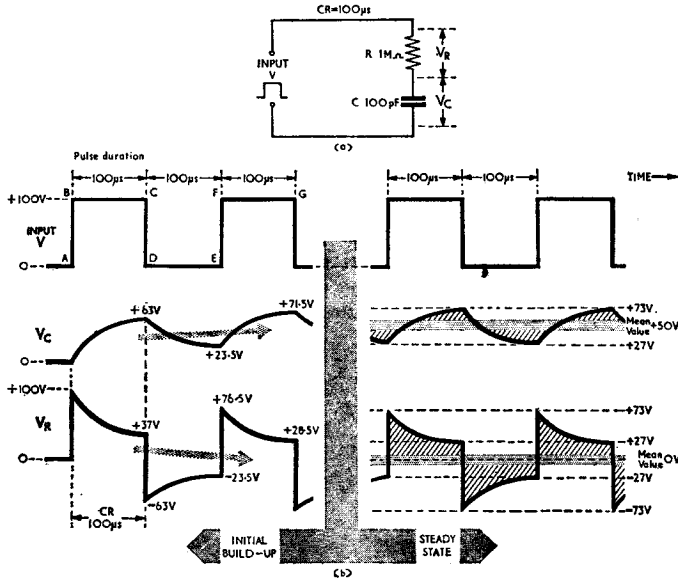


FIG 9. SQUARE WAVE APPLIED TO A MEDIUM CR CIRCUIT

f. F to G. C commences to charge. V_C starts at $+23.5V$ so that the *effective* charging voltage applied to it is $100 - 23.5 = 76.5V$. In time CR ($100\mu s$) V_C therefore rises exponentially from $+23.5V$ to $23.5 + \left(\frac{63}{100} \times 76.5\right) = +71.5V$. V_R falls from $+76.5V$ to $+28.5V$ in the same time ($V_C + V_R = 100V$ at this instant).

The process continues after G until the circuit reaches a state where C is charging and discharging by *equal amounts* on alternate half cycles of the input. After a few cycles V_C settles down about a *mean value* of $+50V$ and V_R about a *mean value* of zero volts. The shaded areas in the waveforms of Fig 9b are equal, indicating that the variations about the mean levels for both V_C and V_R are the same.

Long CR Circuit

Let us assume that a symmetrical square wave of voltage, of amplitude $100V$ and pulse duration $100\mu s$, is applied to a CR circuit where $C = 0.001\mu F$ and $R = 2M\Omega$ (Fig 10a).

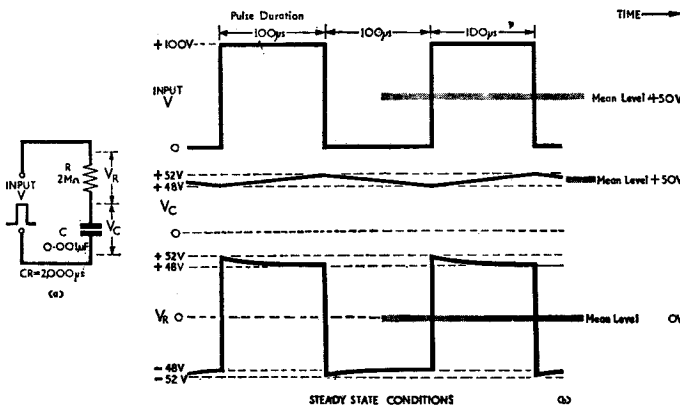


FIG 10. SQUARE WAVE APPLIED TO A LONG CR CIRCUIT

The time constant $CR = 0.001 \times 10^{-6} \times 2 \times 10^6 = 2,000 \mu s$.

This is *more than ten times* the pulse duration so that the combination is a *long CR* to this particular input.

The resultant steady-state waveforms of V_C and V_R are shown in Fig 10b and are derived in much the same way as that for a medium CR circuit. The mean value of V_C *rises* and that of V_R *falls* until eventually the mean value of V_C is $+50V$ and that of V_R is zero volts. In this case however, since we have a long CR , the building-up process takes *longer* and the final swings of V_C about $+50V$ are *much smaller*.

The waveform of V_R is a slightly distorted square wave, almost identical with the input *except for the change of mean level*. The longer CR is made, the less distorted is V_R . Thus for a square wave to be passed by a CR circuit *without distortion* a *long CR* must be used and the output must be taken across the *resistor* (Fig 11).

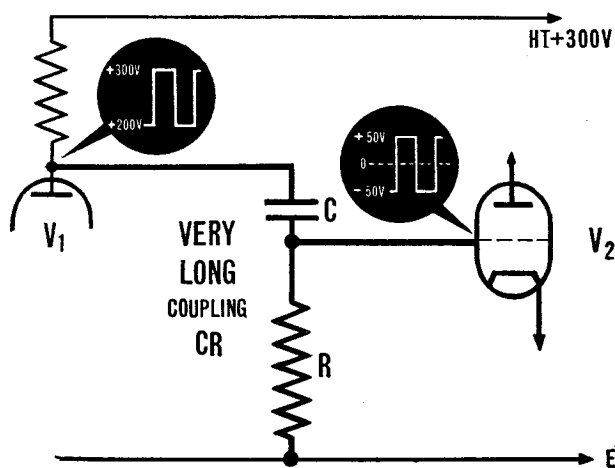


FIG 11. USE OF A LONG CR AS A COUPLING CIRCUIT

Integrating Circuit

Some radar circuits use *triangular* waveforms. One method of obtaining them is to apply a symmetrical square wave to a *long CR* circuit and take the output across the *capacitor*. A CR circuit used for this purpose is called an *integrating circuit* (Fig 12).

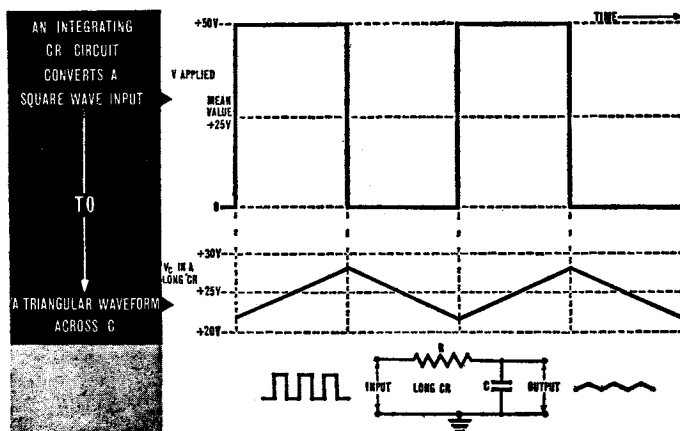


FIG 12. INTEGRATING CIRCUIT

Response of CR Circuits to Asymmetrical Square Waves

The rectangular waveform shown in Fig 13a is one of the more common waveforms found in radar. When such a waveform is applied to a CR circuit the terms short, medium or long CR

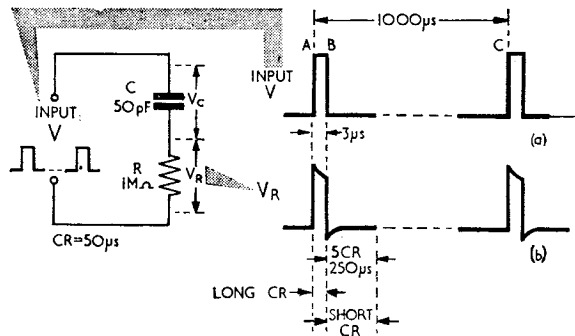


FIG 13. ASYMMETRICAL SQUARE WAVE APPLIED TO A CR CIRCUIT

refer only to the particular portion of the waveform under consideration. In a CR circuit, in which C is 50 pF and R is $1\text{ M}\Omega$, the time constant is:—

$$CR = 50 \times 10^{-12} \times 10^6 = 50 \mu\text{s}.$$

If the asymmetrical square wave shown in Fig 13a is applied to this CR circuit, the circuit acts as a *long CR* to the *short pulse* AB but as a *short CR* to the period BC between pulses. The waveform appearing across R would then be as shown in Fig 13b.

Fig 14 shows the steady-state waveforms of a CR circuit in which the values of C and R are such that the circuit acts as a *long CR* to both parts of each cycle of an asymmetric square wave (two inputs shown). The following points should be noted:—

- C can charge and discharge only by a small amount in each pulse period.
- As the pulse periods are unequal a larger charging current during a short pulse period is needed to balance a smaller discharging current during the long pulse period.

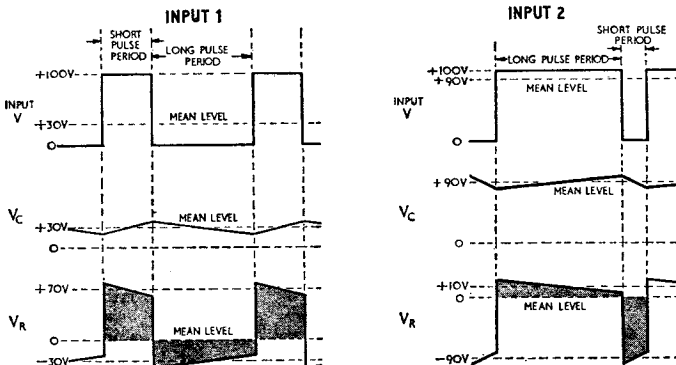


FIG 14. ASYMMETRICAL SQUARE WAVES APPLIED TO A LONG CR CIRCUIT

c. The circuit becomes stable when the shaded areas in Fig 14 are equal. The waveform is then balanced even if it does not *appear* to be so. The mean value of V_R is now zero and that of V_C is equal to the mean value of the input.

d. If the symmetry of the input waveform changes, the top and bottom levels of the V_R waveform change. This may be seen by comparing the waveforms for the two different inputs in Fig 14.

CHAPTER 3

SQUARE WAVES APPLIED TO LR AND LC CIRCUITS

Introduction

We have seen in Part 1 of these notes that a coil of wire has the property of inductance which has the effect of *opposing any change* in the value of current flowing through the coil. It does this by producing a self-induced voltage, or *back e.m.f.*, which is of opposite polarity to the applied voltage.

Circuits which contain inductance L and resistance R are known as LR circuits (Fig 1). When we apply a square wave of voltage to a LR circuit the resultant waveforms of V_L and V_R are similar to the waveforms of V_R and V_C we obtain from CR circuits. LR circuits are not often used *deliberately* in practice in the way that CR circuits are used, because a capacitor is normally needed to 'block' the d.c. component of the resultant waveform. However, every inductor has resistance as well as inductance and so acts as a LR circuit. Allowance must therefore be made for this in the design of radar circuits because of the effect which the application of a square wave to a LR circuit has. Let us see something of this effect.

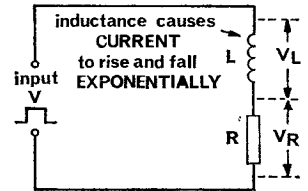


FIG 1. BASIC LR CIRCUIT

Effect of Applying a Square Wave to a LR Circuit

In Fig 2a we have a LR circuit and a simple 'square wave generator', i.e. the battery and the switch. With the switch in the OFF position the voltage V applied to the LR circuit is zero, V_L is zero and so is V_R .

a. *A to B.* When we switch ON, the current cannot rise to its maximum value instantaneously because the back e.m.f. opposes any rise in the current. Initially therefore there is no current in the circuit and the voltage V_R must be zero ($V_R = I \times R$). The *whole* of the applied voltage V thus appears across L and V_L rises *immediately* to the voltage V of the supply, in this case $+100V$; V_R at this instant is zero (Fig 2b).

b. *B to C.* Subsequently the current begins to rise and will continue to rise *exponentially* on a time constant of $\frac{L}{R}$ seconds until, if given time, it reaches its maximum value. Thus V_R rises exponentially and V_L falls at the same rate (at all times $V_R + V_L = V$). After a time equal to $5\frac{L}{R}$ seconds a steady state is reached where $V_R = 100V$ and $V_L = \text{zero}$.

c. *C to D.* If now we switch OFF, the applied voltage V falls immediately from $+100V$ to zero. However, the current in the circuit cannot change instantaneously, because of the effect of the inductance, and the voltage V_R will *remain* initially at $+100V$. Since at all times $V = V_R + V_L$, and the applied voltage V is zero, we have at this instant:

$$0 = V_R + V_L$$

$$\therefore V_L = -V_R = -100V.$$

Thus at the instant of switching OFF, the voltage V_L across L falls *immediately* to $-100V$.

d. *D to E.* Subsequently the current in the circuit falls exponentially on a time constant of $\frac{L}{R}$ seconds, causing V_R to fall exponentially from $+100V$ towards zero and V_L to rise exponentially from $-100V$ towards zero (Fig 2b). Both voltages are zero after $5\frac{L}{R}$ seconds.

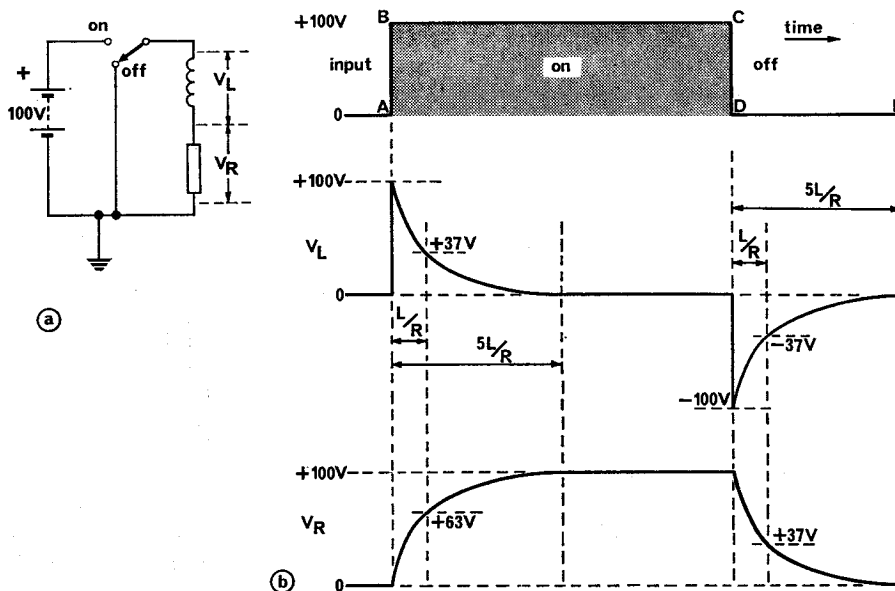


FIG 2. SQUARE WAVE APPLIED TO A LR CIRCUIT

Short, Medium and Long LR Circuits

With LR circuits we compare the time constant L/R seconds against the pulse duration T of the input square wave in the same way that we did with CR circuits. Thus, a short LR circuit is one where the time constant L/R is $0.1T$ (or less); a medium LR circuit is one where L/R is between $0.1T$ and $10T$; and a long LR circuit is one where L/R is greater than $10T$.

The waveforms of V_L and V_R , when a square wave of voltage is applied to short, medium and long LR circuits, are shown in Fig 3. These have the *same shape* as those of CR circuits but the corresponding waveforms appear across *different components*. In general the waveform of V_R in a LR circuit is similar to that of V_C in a CR circuit; V_L in a LR circuit is similar to V_R in a CR circuit. For example, positive- and negative-going pips appear across the *resistor* of a short CR and across the *inductor* of a short LR circuit. (Compare the short LR waveforms of Fig 3 opposite with those for the short CR circuit of Fig 7 on p 71).

Note that LR circuits may be used as differentiating and integrating circuits in much the same way that CR circuits are used. A *short LR* circuit in which the output is taken across the *inductor* acts as a *differentiating circuit*. A *long LR* circuit in which the output is taken across the *resistor* acts as an *integrating circuit*.

Effect of Applying a Square Wave to a Parallel LC Circuit

Many radar circuits contain conventional parallel tuned circuits of the type discussed in Part 1 of these notes. These tuned circuits may form the anode load of a valve or the collector load of a transistor and, since the input to the stage may be a pulse of voltage, it is important to discuss the effect of applying a square wave to such a circuit.

Fig 4a shows a parallel tuned circuit connected as the anode load of a valve which is being cut on and off by a switching square wave at the grid. When the valve is cut on, the supply current is steady at a value I_a determined by the circuit constants; when the valve is cut off, the supply current is zero (see Fig 4b).

SQUARE WAVES APPLIED TO LR AND LC CIRCUITS

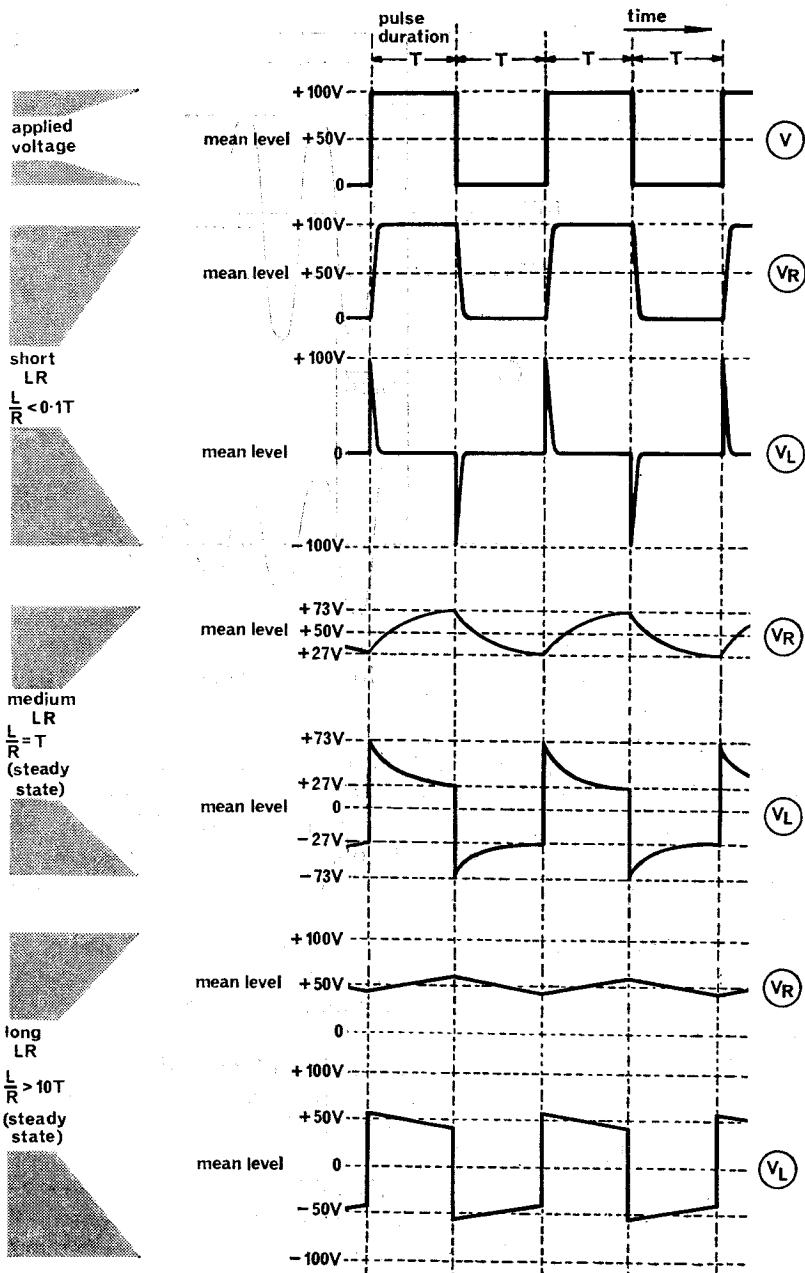


FIG 3. SQUARE WAVE APPLIED TO SHORT, MEDIUM AND LONG LR CIRCUITS, WAVEFORMS

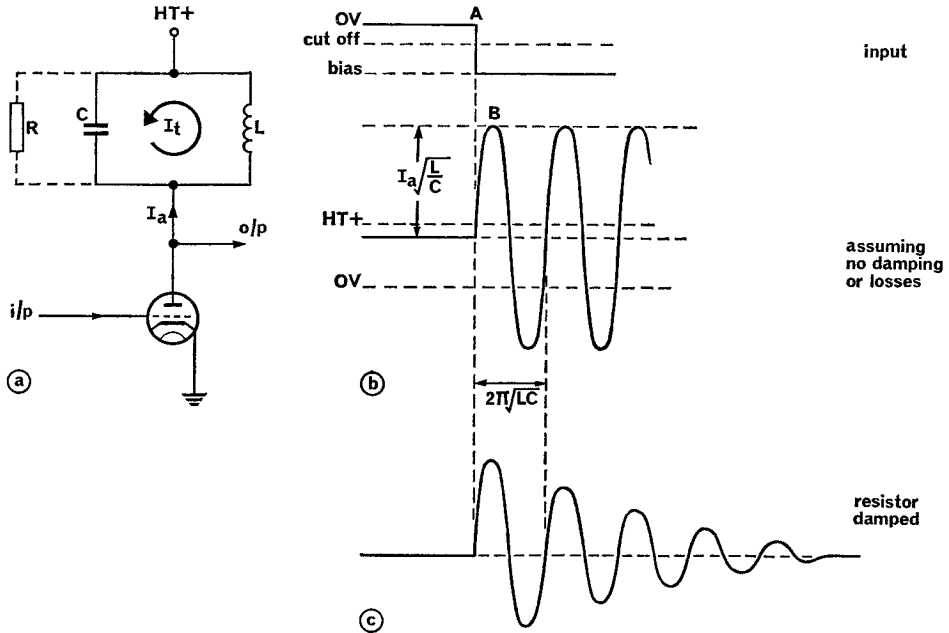


FIG. 4 TUNED CIRCUIT WAVEFORMS

Assuming that the valve has been switched on for some time, the whole of the anode current I_a will be flowing through the inductance L . The anode voltage will be at approximately HT , as the only losses are due to the small built-in resistance of the inductor.

At instant A the valve is suddenly cut off and the supply of current ceases. However, the current flowing in an inductance cannot change instantaneously so the current I_a will continue to flow in the tuned circuit following the path shown as I_t in Fig 4a. This causes the capacitor C to charge up, and the voltage on C will begin to oppose the flow of current I_t . After a time depending on the tuned circuit constants, current flow ceases and the capacitor is fully charged. (Point B in Fig 4b).

The voltage on the capacitor now causes the current I_t to build up in the opposite direction through the inductor, until maximum current I_a is again flowing and the voltage on the capacitor is again zero. The oscillatory action will then carry on at the tuned circuit resonant frequency as there are assumed to be no losses.

What this really means is that at the instant of switch off we had a given amount of energy stored in the inductor ($\frac{1}{2}LI^2$): this energy was then transferred to the capacitor ($\frac{1}{2}CV^2$), then back to the inductor and so on. The rate of transfer is such that the oscillating output voltage is at the frequency of the tuned circuit $1/2\pi\sqrt{LC}$.

The output voltage depends on the value of the initial current flowing in the circuit, but it also depends on the L/C ratio and quite high voltages can be obtained from a circuit arrangement of this type. Two applications are worth mentioning, the ignition system of a motor car and a circuit called the ringing oscillator which is often used to produce high voltage trigger pulses in radar.

Figure 4c shows the effect on the output of having a resistance in parallel with the tuned circuit and thus represents the practical case where the circuit is loaded by the following stage. As can be seen the amplitude of the output waveform decreases and the circuit is said to be damped. The lower the value of resistance in parallel with the tuned circuit, the faster the oscillations die away.

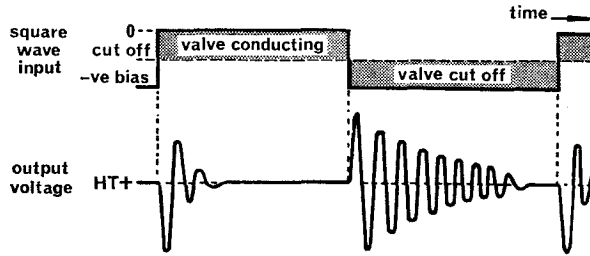


FIG. 5 DAMPED OSCILLATIONS IN A RINGING OSCILLATOR

We can now consider what happens in the circuit when the valve is switched on. The resistance of the conducting valve is effectively in parallel with the tuned circuit and causes these oscillations to die away fairly rapidly, giving the waveforms shown in figure 5.

In some circuits it is necessary to protect a transistor from high voltage rings as the circuit is switched off. This is often done by connecting a diode across the tuned circuit with its cathode connected to the HT+ line. As soon as the anode rises above HT+, the diode conducts and introduces losses in the tuned circuit. This prevents very high voltages being developed and causes the ringing to die away extremely quickly.

CHAPTER 4

LIMITING CIRCUITS

Introduction

In radar it is often necessary to *remove* the positive or negative extremes of an input signal. This may be done to *improve the shape* or to '*clip*' unwanted parts of an input waveform. A circuit which does this is known as a *limiting circuit* or *limiter*; an alternative name is a *clipping circuit* (Fig 1).

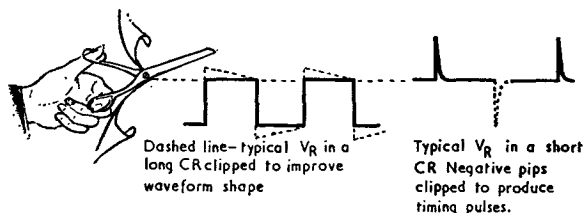


FIG 1. APPLICATIONS OF LIMITING CIRCUITS

Diode Limiters

A diode limiter is the simplest form of limiting circuit and consists of a diode and a single fixed resistor. There are two types of diode limiter—*series* and *parallel* limiters. These in turn are each sub-divided into *positive* and *negative* limiters. Thus we have four distinct circuits (Fig 2). The diode may be either a thermionic valve or a semiconductor type.

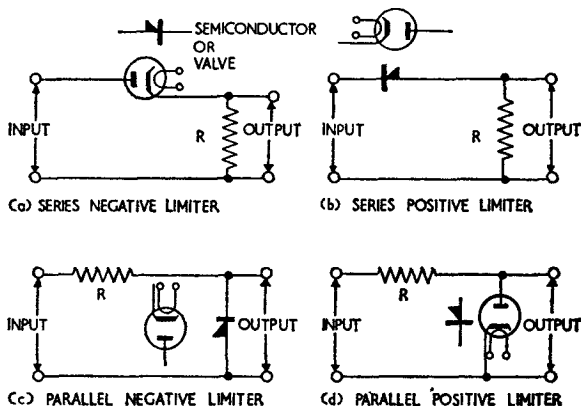


FIG 2. DIODE LIMITERS

Action of Diode Limiter

The resistance R_D between the anode and cathode of a diode is *very high* (say $10M\Omega$) when the anode is negative with respect to its cathode (diode cut off). However when the anode is made positive with respect to its cathode the diode conducts and R_D then falls to a *low value* (say $1k\Omega$) (see Fig 3).

Thus if we apply an alternating voltage V of peak value $100V$ to any of the limiter circuits shown in Fig 2, each circuit operates as a voltage divider consisting of a fixed resistor R , normally about $100k\Omega$, in series with the diode resistance R_D which is either *very low* (when diode conduct-

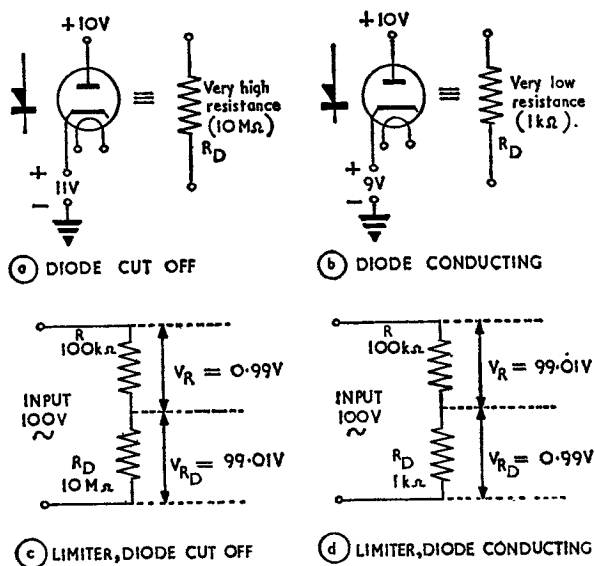


FIG 3. ACTION OF DIODE LIMITER

ing) or *very high* (when diode cut off). Therefore when the input polarity is such that the diode is cut off the voltage developed across R_D is much larger than that developed across R . For the figures given, the peak voltage across the diode is:—

$$\frac{R_D}{R_D + R} \times V = \frac{10M\Omega}{10M\Omega + 100k\Omega} \times 100 = 99.01V.$$

The peak voltage across the resistor is then 0.99V (see Fig 3c).

When the input polarity is such that the diode is conducting, the voltage developed across R is much larger than that developed across R_D . For the figures given, the peak voltage across R is:—

$$\frac{R}{R + R_D} \times V = \frac{100k\Omega}{100k\Omega + 1k\Omega} \times 100 = 99.01V.$$

The peak voltage across the diode is then 0.99V (see Fig 3d).

Thus practically the whole of the applied voltage is developed across either the resistor or the diode depending upon whether the valve is conducting or cut off.

Series Limiters

In a series limiter the diode is connected *in series* with the circuit to which we apply the output waveform, and the output voltage is taken across the *resistor*. Fig 4 shows the two forms of series limiter.

If the input to the series *negative* limiter shown in Fig 4a is a sine wave which varies about earth the diode conducts only during the positive-going half-cycles and is cut off during the negative-going half-cycles. When the diode is conducting the voltage developed across it is negligible and the output voltage V_R almost equals the input voltage V_{in} . When the diode is cut off V_R is practically zero. This circuit therefore clips the portion of the input waveform which goes *negative* with respect to earth.

If we turn the diode round we have a series *positive* limiter as shown in Fig 4b. In this circuit the diode can conduct only during the negative-going half-cycles and so the *positive-going* portion of the input waveform is clipped.

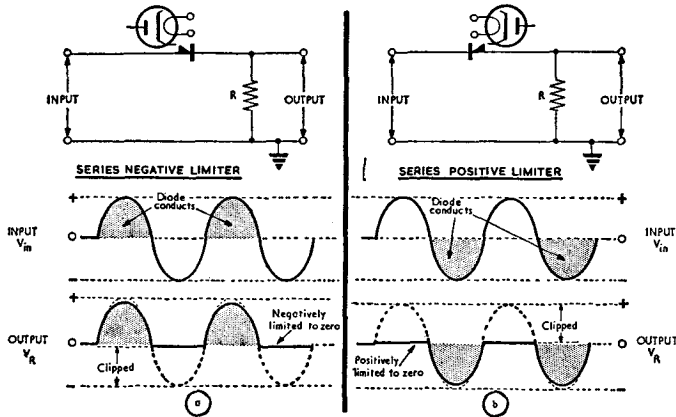


FIG 4. SERIES DIODE LIMITERS, CIRCUITS AND WAVEFORMS

Parallel Limiters

In a parallel limiter the diode is connected *in parallel* with the component to which we apply the output waveform and the output voltage is taken across the *diode*.

If the input to the parallel positive limiter shown in Fig 5a is a sine wave which varies about earth the diode conducts only during the positive-going half-cycles and the voltage developed across it (the output voltage V_{out}) is then practically zero. When the diode is cut off on the negative-going half-cycles of input practically the whole of the input voltage is developed across the diode and applied to the output terminals. This circuit therefore clips the portion of the input waveform which goes *positive* with respect to earth.

If we wish to remove the negative-going portion of the input waveform we simply turn the diode round to give the parallel *negative* limiter shown in Fig 5b.

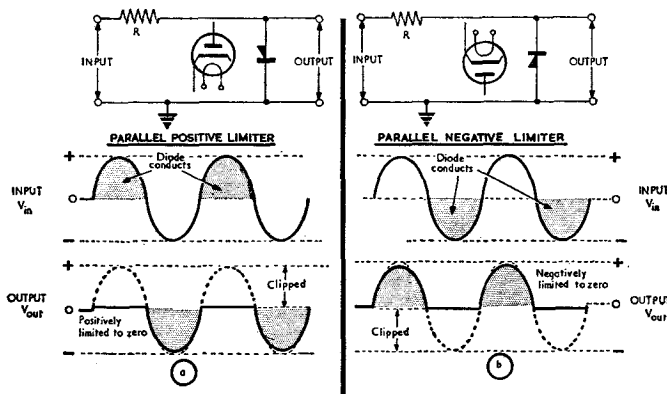


FIG 5. PARALLEL DIODE LIMITERS, CIRCUITS AND WAVEFORMS

Limiting to Voltages other than Zero

The circuits we have just considered *all limit to zero volts*, i.e. they use earth as a reference and cut off the portion of the input waveform which rises above, or falls below, zero volts.

In practice we often need to clip the portion above or below some reference voltage *other than zero*. This may be done by using slightly modified versions of the basic limiting circuit

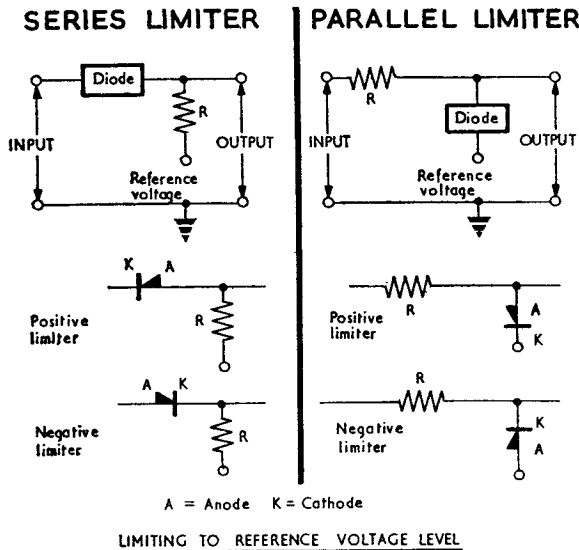


FIG 6. LIMITING TO VOLTAGES OTHER THAN ZERO

(Fig 6). The waveform may be limited to any positive or negative reference level by holding the appropriate electrode of the diode at the required value of bias voltage. Whether the waveform is positively limited or negatively limited to this reference value depends upon how the diode is connected.

Fig 7 shows examples of biased diode parallel limiters. Series limiters may be shown in a similar manner.

In Fig 7a, with a positive reference bias applied to the cathode, the diode is cut off until the input voltage rises *above* the bias level. When the diode is cut off, practically the *whole* of the input voltage is developed across it and appears as V_{out} . When the diode is conducting the voltage developed across it is practically *zero* and V_{out} is then limited to the positive bias level E . We therefore have *positive limiting above* the bias level. The waveforms of the other circuits in Fig 7 may be deduced in a similar manner.

Although bias batteries have been shown in each of the previous examples, in practice the reference bias voltage is usually obtained from a more convenient source, e.g. from a potential divider connected across a supply network. In addition, although all the inputs so far have been shown as sine waves any waveform may be used.

A typical circuit, which gives *negative limiting* to a *positive* bias level, is shown in Fig 8. This circuit is using a series *negative* limiter with the resistor connected to a reference bias voltage of $+30V$. Because of this, the diode will conduct only when the input voltage rises above $+30V$. If a square wave input of $50V$ amplitude is applied to a short CR differentiating circuit the input waveform to the limiter is a series of positive- and negative-going pips each of $50V$ amplitude as shown in Fig 8. The limiter diode conducts only on the peaks of the positive-going pips when the input voltage is *above* the reference bias level. Thus the output from the limiter is a series of *very narrow* pulses of $20V$ amplitude, *positive-going* from $+30V$. These pulses may be used as timing or triggering pulses.

Combined Positive and Negative Limiting

Combined positive and negative limiting may be obtained from one circuit using two diodes arranged as shown in Fig 9. Each diode conducts in turn so that both half-cycles of the input

FIG 7. BIASED DIODE PARALLEL LIMITERS, CIRCUITS AND WAVEFORMS

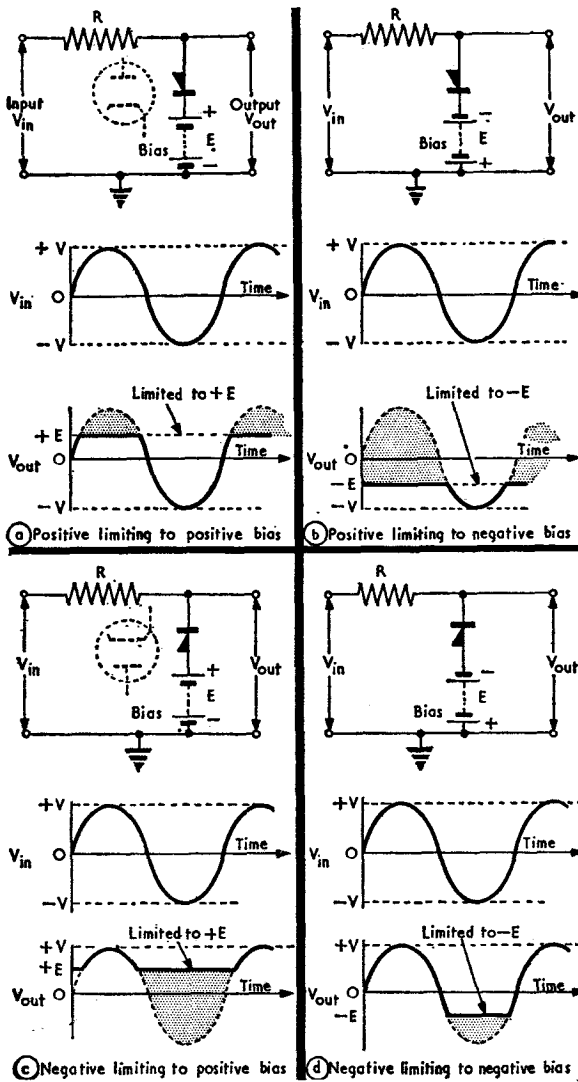
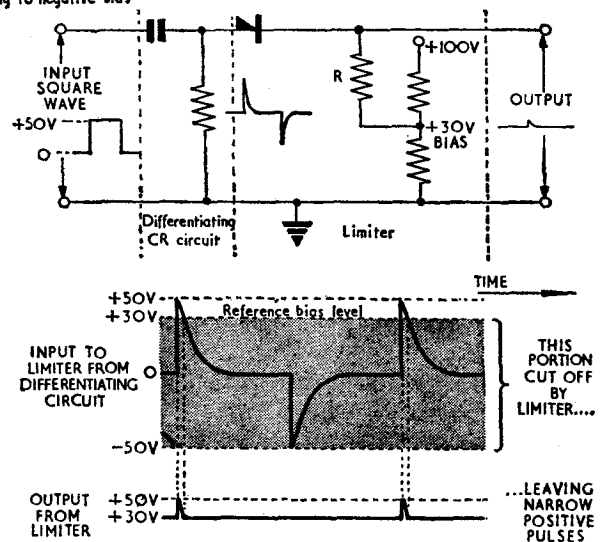


FIG 8. PRACTICAL APPLICATION OF DIODE LIMITER



are limited. By making the reference bias voltages applied to the diodes equal, an approximate square wave output is obtained from a sine wave input. A double diode-limiter used for this purpose is sometimes called a *diode squarer*. The output amplitude is of course *smaller* than that of the input.

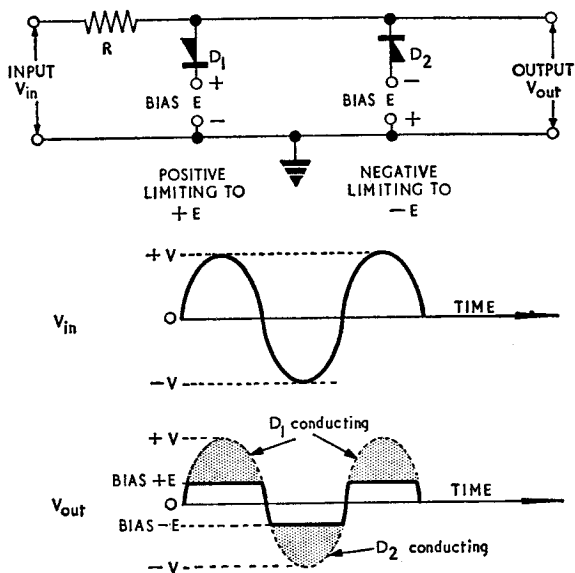


FIG 9. DIODE SQUARER, CIRCUIT AND WAVEFORMS

Limiting in a Triode Amplifier

There are two ways in which a triode amplifier may be used as a limiting circuit:—

a. The input waveform may be made large enough to drive the valve *beyond cut-off* at the peaks of the negative half-cycle. The anode current I_a is then zero and the anode voltage V_a remains at h.t. + whilst the valve is cut off (Fig 10a). *Grid cut-off limiting* gives *negative* limiting below cut-off at the grid of the valve.

b. A resistor of high value may be inserted in the grid-cathode path of the valve. The grid and cathode act as a diode, with the triode grid functioning as the diode 'anode'. Thus the grid-cathode circuit with its large series resistor (known as a *grid stopper*) forms an effective diode parallel limiter (Fig 10b). If the grid is driven positive by the input signal, grid current flows and the grid-cathode path of the valve has a very low resistance, of the same order as that of the conducting diode in a diode limiter. Thus most of the input is developed across the grid stopper and the grid-cathode voltage V_g is practically zero. Conversely when the grid goes negative the effective grid-cathode resistance is very much higher than that of the grid stopper and most of the input voltage appears as V_g . This action is known as *grid current limiting* and gives *positive* limiting above zero volts at the grid of the valve.

Both processes may be used together to give combined negative (grid cut-off) and positive (grid current) limiting. Such an arrangement may be used to convert a sine wave input to an approximate square wave output. When used in this manner a triode limiter is referred to as a *squarer*. A triode squarer produces a better square wave output than that from a diode squarer because of the amplifying properties of the triode.

A typical triode squarer circuit and its associated waveforms are illustrated in Fig 11. During the positive half-cycle of the input, grid current limiting maintains V_g practically at zero

FIG 10. TRIODE LIMITER, CIRCUIT AND WAVEFORMS

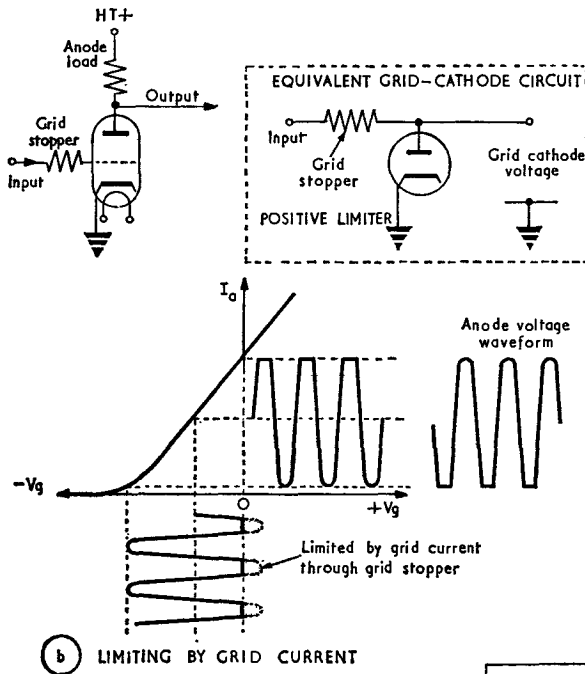
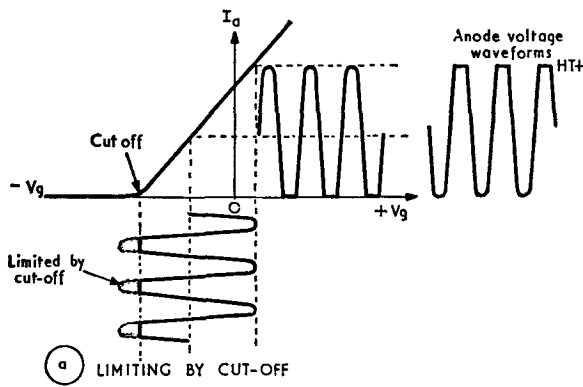
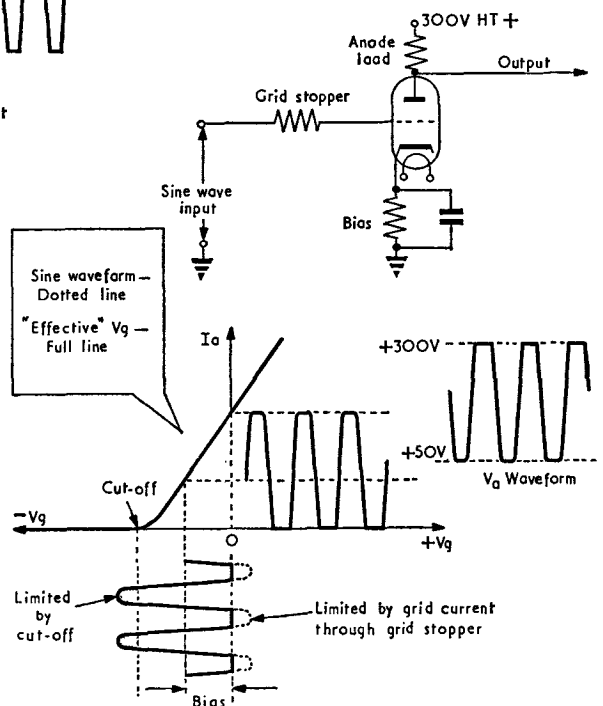


FIG 11. TRIODE SQUARER, CIRCUIT AND WAVEFORMS



volts and the output voltage V_a is at some value determined by the circuit constants (say $+50V$). On the negative-going half-cycle of the input, when V_g falls below cut-off, the anode voltage V_a rises to h.t. $+$ (say $+300V$) and remains there until V_g again rises above cut-off. We can see from Fig 11 that the output waveform is not perfectly square: it varies slightly while the valve is conducting and takes a short but definite time to rise and fall. The waveform may be improved by applying the squarer output to a diode limiter, arranged to clip the curved portion of the output waveform.

The frequency of the output is exactly the same as that of the sine wave input and the waveform is normally sufficiently square to be converted into triggering pulses which can then be used to synchronize other stages in the equipment.

Limiting in a Pentode Amplifier

In a normal valve amplifier circuit with a resistive anode load, if V_g rises, I_a rises and the increased voltage drop across the load causes V_a to fall. Fig 12a shows a family of typical pentode I_a - V_a curves. Let us see the effect of using a *very large value resistor* as the anode load R_L . We shall assume that R_L is so large that when V_g is zero (curve B) the anode current is such that the voltage drop across R_L brings V_a down to only $50V$. From Fig 12a we can see that when V_g is zero and V_a is $50V$, I_a is $10mA$. However the coincident values of $I_a = 10mA$ and $V_a = 50V$ apply to other curves also, i.e. curve C ($V_g = -2V$) and curve A ($V_g = +2V$). This means that if we vary V_g positively *above* $V_g = -2V$, there is *no change* in I_a and so V_a *remains*

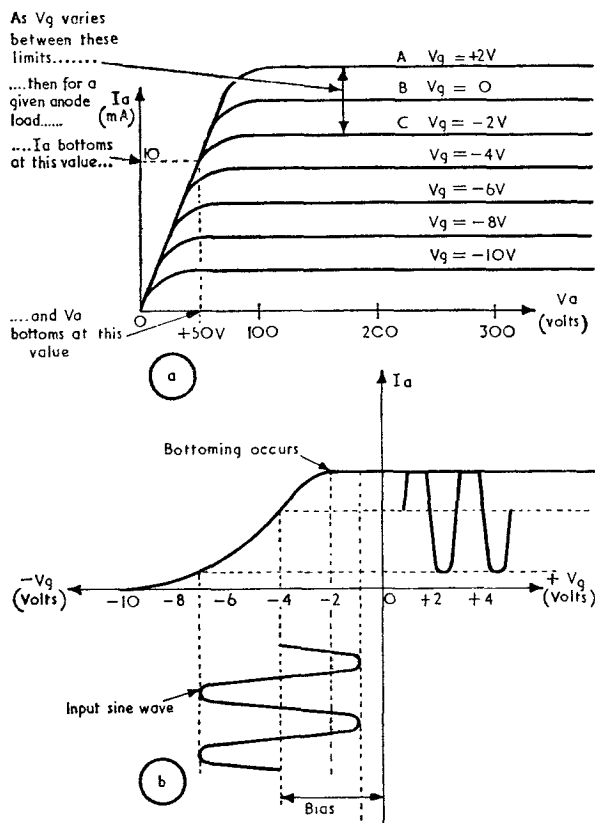


FIG 12. BOTTOMING IN A PENTODE

at 50V. This condition where a change of V_g does not cause a change in either I_a or V_a is known as *anode bottoming*. It results from the overlap of the curves below the knee of the characteristic at large values of anode load. The effect on the I_a - V_g curve is shown in Fig 12b. In this example, anode bottoming occurs at $V_g = -2V$. However by increasing the value of anode load resistor R_L , bottoming at even greater negative values of V_g can be obtained.

A pentode amplifier may be used as a *squarer* to produce square waves from a sine wave input by limiting *both* the positive and the negative half-cycles of the input. The *positive* half-cycle is limited by *bottoming* and the *negative* half-cycle by *cut-off* (Fig 13). Although the grid

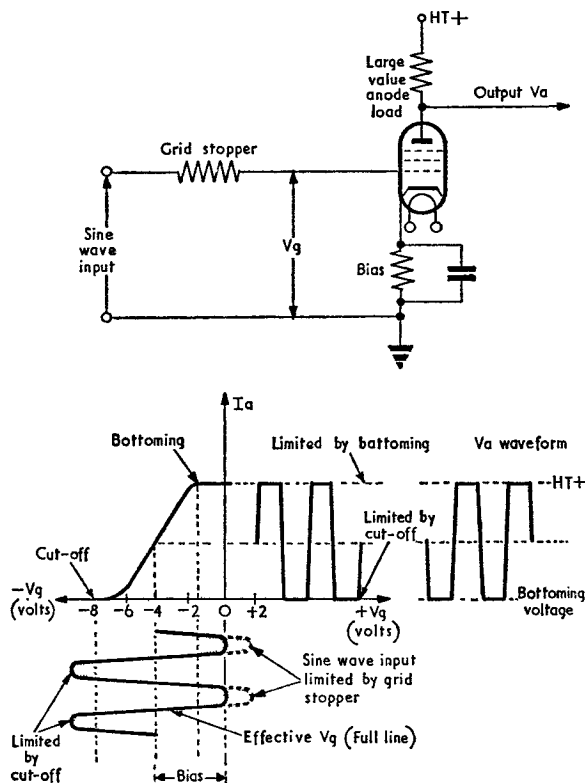


FIG 13. PENTODE SQUARER, CIRCUIT AND WAVEFORMS

stopper in a pentode squarer does not usually have any effect on the output waveform it is still retained to limit the positive swings of the V_g waveform. By limiting the grid current the input impedance of the amplifier is kept at a high value.

The output from a pentode squarer is a better square wave than that from a triode squarer. The top and bottom of the waveform are perfectly flat. In addition, a pentode with a *short grid base* can be selected and this, coupled with the high gain of a pentode, gives a *steep-sided* output waveform. It can be made more steep if a very large amplitude input sine wave is used so that the 'effective' portion of the input has almost vertical edges. For this reason inputs of 100V peak are common.

Symmetry of Output from Squarers

The *symmetry* of the square wave output from triode and pentode squarers depends upon the operating point selected for a given valve. This may be adjusted by varying the *bias*. When the

valve is operating under Class A conditions the output is *symmetrical*. The effect of varying the bias either side of the Class A biasing point is illustrated in Fig 14. The output is now *asymmetrical*. Note however that the slope of a sine wave decreases towards the peaks. Hence the use of a large bias to produce narrow pulses results in deterioration in the steepness of the edges of the square wave output.

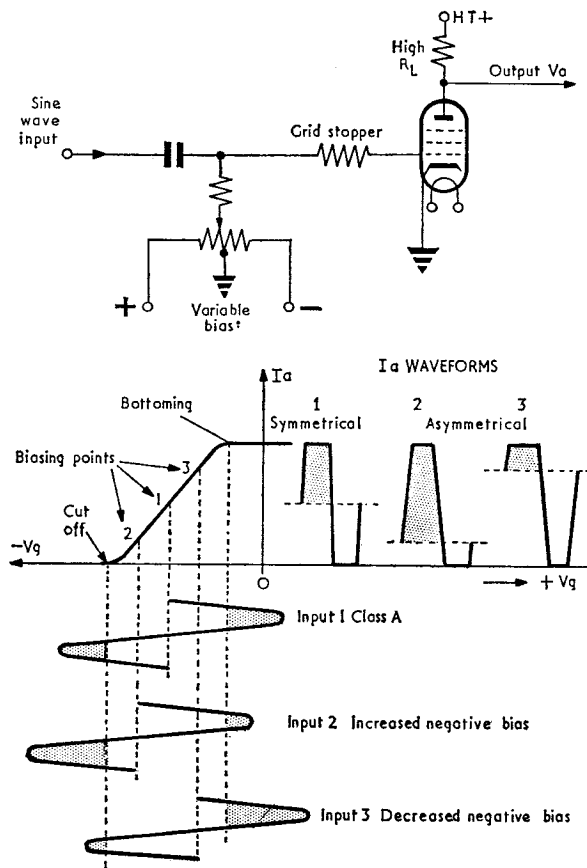


FIG 14. VARIATION OF SYMMETRY IN PENTODE SQUARER

Cathode Follower Limiting

The cathode follower circuit is considered in some detail in Part 1 of these notes. It is an amplifier in which the output is taken across the *cathode* load resistor, the phase of the output being the *same* as that of the input. Since the cathode resistor is not decoupled, 100 per cent voltage negative feedback is applied between output and input. This has two main effects:

- The voltage gain of the amplifier is always *less* than unity, i.e. output smaller than input.
- The input impedance is *very high* and the output impedance *very low*. It therefore provides a means of *matching*.

Limiting in a cathode follower (Fig 15) is similar to that in a triode amplifier: negative limiting is achieved by grid cut-off and positive limiting by grid current flow through a grid

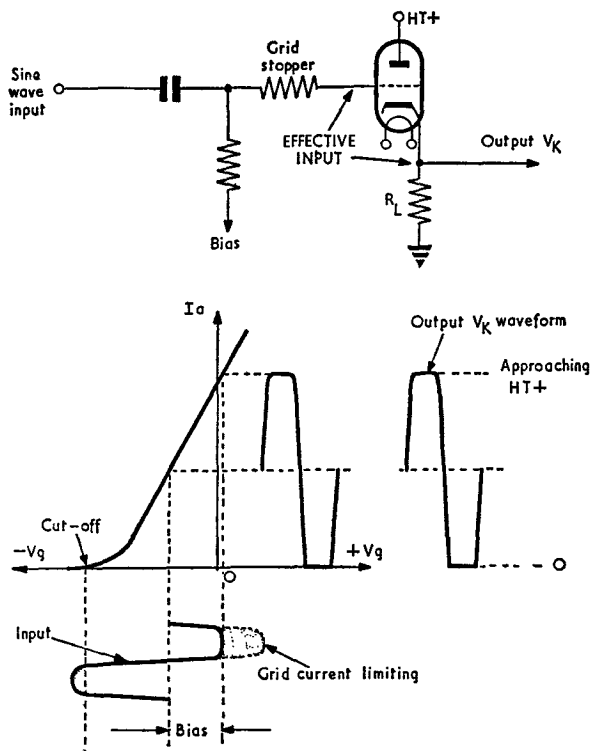


FIG 15. CATHODE FOLLOWER LIMITER, CIRCUIT AND WAVEFORMS

stopper. However in a cathode follower the *effective* grid-cathode input is the *difference* between the input and output voltages. Hence to provide the limiting action a *large* amplitude input is required (typically 100V peak).

Transistor Limiting

A transistor may be used as a squarer by applying a large amplitude sine wave to its base. The basic circuit of a p-n-p transistor limiter is shown in Fig 16a and the resulting waveforms in Fig 16b.

It will be remembered that the bias conditions for a p-n-p transistor are *opposite* to those for a valve, i.e. the application of a *positive* bias to the base *reduces* the base-collector current and if the base voltage is sufficiently positive the transistor will cut off. Thus the positive-going half-cycles of the input sine wave are limited by transistor cut-off, the output voltage V_C at the collector then being equal to the supply voltage, in this case — 18V. On the negative-going half-cycles of the input, if the collector load is sufficiently large, e.g. 100k Ω , the transistor *bottoms* in much the same way as a pentode, because of the similarity of their characteristics. When this happens the output voltage V_C at the collector rises almost to earth potential (zero volts). An approximate square wave output then results. The *symmetry* of the square wave output may be adjusted by altering the bias applied to the base.

The purpose of the limiting resistor in the base is much the same as that of a grid stopper in a pentode, i.e. it limits the base current and prevents damage to the transistor.

Summary

Limiting circuits of the type described in this chapter have many uses in radar. As already noted they may be used to improve the shape of a pulse, they may be used to produce accurate

timing pulses from differentiated square wave inputs, or to produce square waves from sine wave inputs. These, and other, applications of limiting circuits will be considered at appropriate parts of these notes.

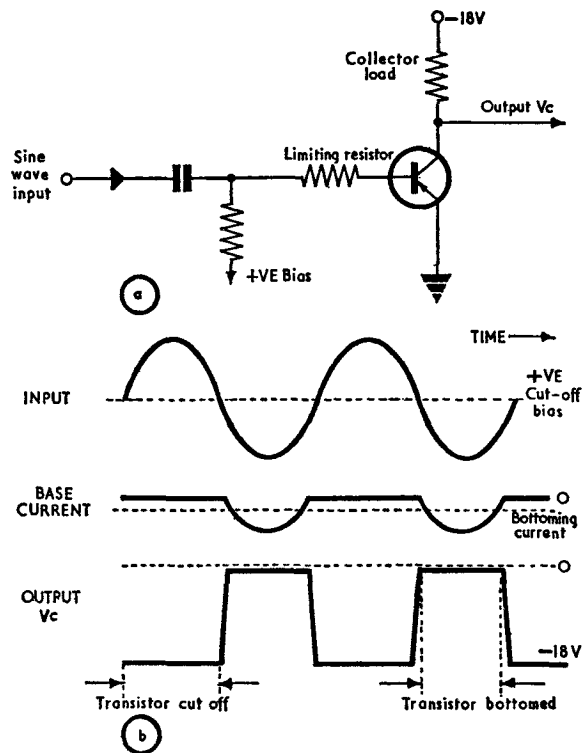


FIG 16. TRANSISTOR LIMITER, CIRCUIT AND WAVE FORMS

CHAPTER 5

CLAMPING CIRCUITS

Introduction

We have seen that limiters prevent a waveform from rising above or falling below a pre-determined voltage by *cutting off* part of the waveform. However in radar there is also a need to *change the reference level* of a waveform *without reducing its amplitude*. Circuits which move waveforms 'up' or 'down' in this way are known as *clamping circuits* because their effect is to fix or *clamp* the top or bottom level of the waveform to a required voltage. The difference between limiters and clamping circuits is illustrated in Fig 1.

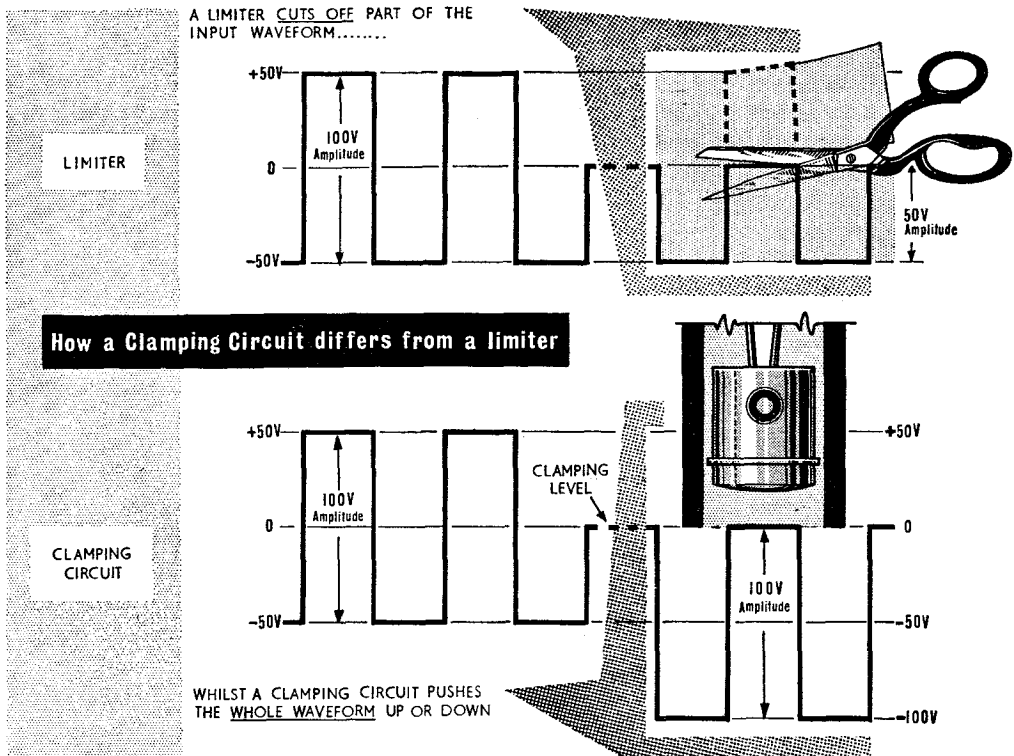


FIG 1. ESSENTIAL DIFFERENCES BETWEEN LIMITING AND CLAMPING

When a square wave is applied to a CR circuit the voltage developed across the resistor is such that its *mean value* is always zero, the d.c. component of the input being 'blocked' by the capacitor (Fig 2a). If we take the voltage across the resistor and apply it as the input to a clamping circuit any of the waveforms shown in Fig 2b (and many more) may be obtained. The waveform may be made to vary with respect to *any* reference level, *its amplitude remaining unchanged*.

An alternative name for clamping circuits is 'd.c. restorers'. This name originates from the fact that any waveform, after passing through a capacitor, has its d.c. component removed and it is often necessary to 'restore' the d.c. component of the input signal in the output waveform.

This restoration of the d.c. component is done by clamping circuits (d.c. restorers) although, as we have seen, the clamping circuit may be adjusted such that the output waveform varies about *any* selected level, not merely the d.c. level of the input signal.

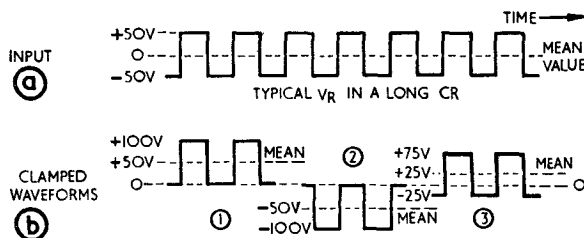


FIG 2. 'RESTORATION' OF DC VOLTAGE BY A CLAMPING CIRCUIT

It may be thought that the necessary d.c. component could be added by using a bias circuit. A *fixed* d.c. voltage is not suitable however because it is necessary to arrange for the d.c. component to alter automatically as the amplitude or pulse duty factor of the input alters. As we shall see later a clamping circuit will give the required variation.

Basic Clamping Circuit

A clamping circuit consists simply of a *long CR* circuit in which the resistor is shunted by a thermionic or semiconductor diode. The direction of connection of the diode in the circuit decides the direction of clamping. Fig 3 shows a basic clamping circuit and also the conditions existing when the diode is cut off and when it is conducting. When the diode is cut off, R is shunted by the very high impedance of the non-conducting diode and the effective resistance is

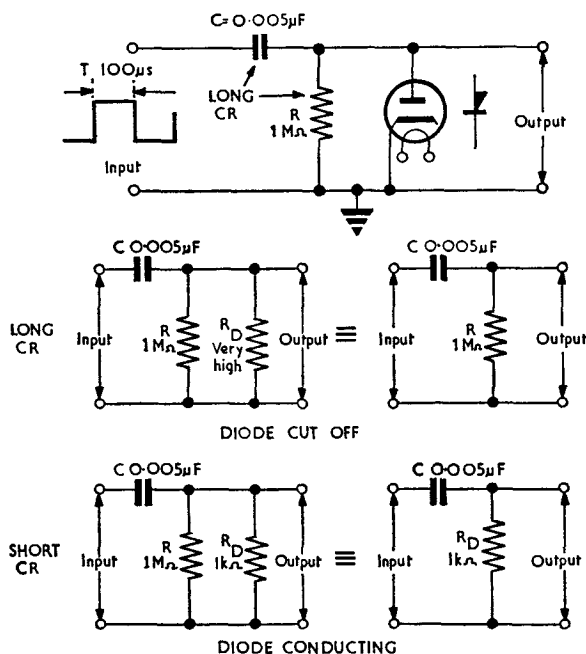


FIG 3. BASIC ACTION OF DIODE CLAMP

R , in this example $1M\Omega$. For a symmetrical square wave input of pulse duration $100\ \mu s$ the circuit now acts as a *long CR*. The time constant is:

$$CR = 0.005 \times 10^{-6} \times 10^6 = 5,000\ \mu s.$$

This is greater than *ten times* the pulse duration.

When the diode is conducting, R is shunted by the very low impedance of the conducting diode and the effective resistance is merely that of the diode R_D , in this example $1k\Omega$. For the *same* square wave input the circuit now acts as a *short CR* where the time constant is:

$$CR_D = 0.005 \times 10^{-6} \times 10^3 = 5\ \mu s.$$

This is less than *one-tenth* of the pulse duration of the input.

Thus a clamping circuit acts either as a *long CR* or as a *short CR* depending upon whether the diode is cut off or conducting.

Positive Clamping to Zero Volts

Fig 4a illustrates a positive clamping circuit. In this the anode of the diode is earthed and the *bottom* of the output waveform is clamped to *zero volts*. The output waveform thus varies between zero and a *positive* value (Fig 4b). Note that since R and the diode are in parallel the

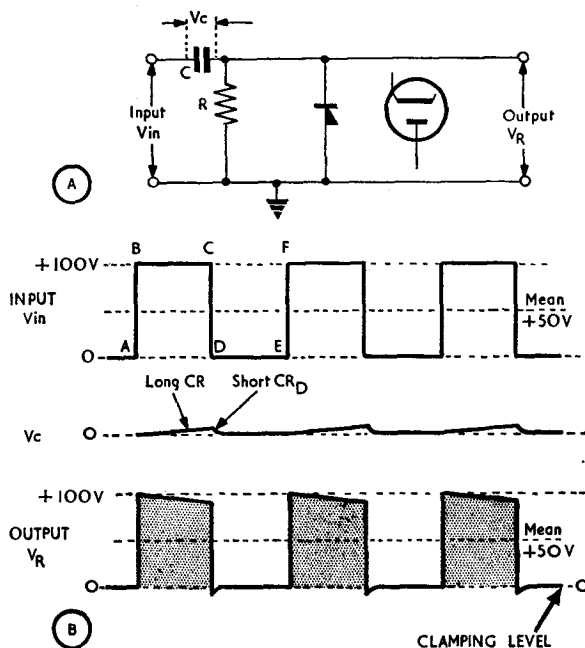


FIG 4. POSITIVE CLAMPING TO ZERO VOLTS, POSITIVE-GOING INPUT

output voltage always equals the voltage V_R developed across R . In addition, as in any CR circuit, the input voltage $V_{in} = V_C + V_R$ at all times. The waveforms of Fig 4b are obtained as follows:—

A to B. The input rises by 100V from zero. The capacitor C is initially uncharged and cannot change its charge immediately. Hence V_R rises *instantly* to + 100V and since this voltage is applied to the cathode of the diode, the anode being earthed, the diode is cut off
B to C. With the diode cut off C charges on a *long* time constant of CR seconds and V_C rises by a *small* amount. Thus V_R falls by the same amount.

C to D. The input falls by 100V to zero and since V_C cannot change immediately V_R also falls by 100V to a small *negative* potential. This causes the diode to *cut on*.

D to E. With the diode conducting, C discharges on a *short* time constant of CR_D seconds. Both V_C and V_R quickly return to zero volts and the diode is cut off.

E to F. The input rises again by 100V from zero and the cycle is repeated.

Except for the small negative pips the output V_R is clamped to a base level of *zero volts* and is *positive-going* from this level. DC restoration has taken place and the *mean level* of the output equals that of the input. It will be remembered that the mean value of voltage for V_R in a long CR in the absence of clamping is *zero*.

For a *negative-going* square wave input the action is as follows (Fig 5):

A to B. The input falls by 100V from zero and, since V_C cannot change instantly, V_R falls to $-100V$. This causes the diode to *conduct*.

B to C. With the diode conducting, C charges on a *short* time constant of CR_D seconds and V_C quickly falls to $-100V$. V_R returns to zero in the same time and the diode is *cut off*.

C to D. The input rises by 100V to zero. V_C cannot change instantly and remains at $-100V$ so V_R rises by 100V from zero to $+100V$. The diode remains cut off.

D to E. With the diode cut off, C discharges on a *long* CR towards zero volts but rises by only a *small* amount in the time available. V_R falls by the same amount.

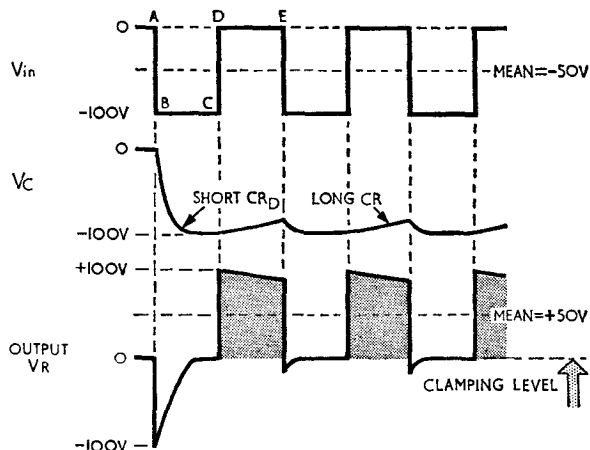


FIG 5. POSITIVE CLAMPING TO ZERO, NEGATIVE-GOING INPUT

Thereafter the action is as described for the positive-going input. Apart from the initial negative spike in Fig 5, the output V_R is the same in both cases. In Fig 5, the mean d.c. level of $-50V$ is restored to $+50V$ in the output. No matter what values the input varies between, the output from the circuit described is *clamped to zero volts* and rises *positively* by an amount equal to the amplitude of the input.

Negative Clamping to Zero Volts

Fig 6a illustrates a negative clamping circuit. In this the *cathode* of the diode is earthed and the *top* of the output waveform is clamped to *zero volts*. The output waveform thus varies between zero and a *negative* value (Fig 6b). The waveforms of Fig 6b are obtained as follows:

A to B. The input rises by 50V from zero. C cannot change its charge instantly so V_C remains at zero volts and V_R rises to $+50V$, cutting on the diode.

B to C. With the diode conducting, C charges on a *short* time constant of CR_D seconds. V_C rises quickly to $+50V$ and V_R falls to zero in the same time. The diode is now cut off.

C to D. The input falls by $100V$ from $+50V$ to $-50V$. Since V_C cannot change instantly V_R falls by $100V$ to $-100V$, keeping the diode cut off.

D to E. With the diode cut off, C discharges on a *long* time constant of CR seconds causing V_C to fall *slightly* from $+50V$ and V_R to rise slightly from $-100V$.

E to F. The input rises by $100V$ from $-50V$ to $+50V$. V_C cannot change instantly so V_R rises by $100V$ to a slightly *positive* value, cutting on the diode.

F to G. With the diode conducting, C charges on a *short* CR_D and V_C rises quickly to $+50V$ as V_R returns to zero, producing the familiar pip. The diode is now cut off.

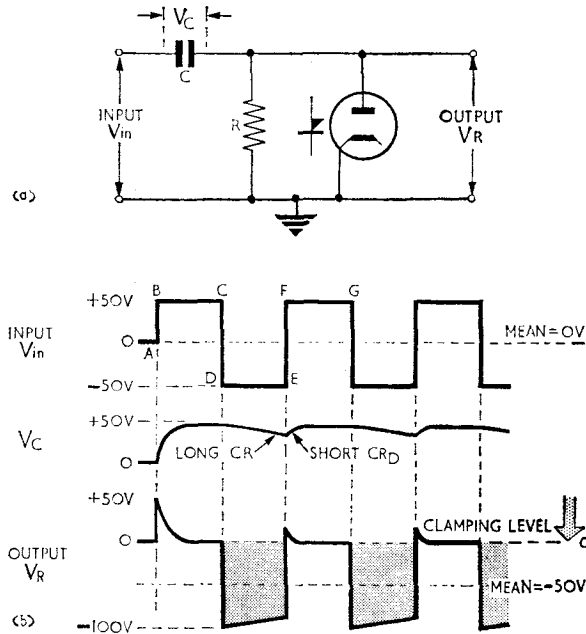


FIG 6. NEGATIVE DIODE CLAMP, CIRCUIT AND WAVEFORMS

Thereafter the action is repeated. Except for the small positive pips the output V_R is *clamped to zero volts*, the output varying *negatively* from this level. DC restoration has also taken place, as a comparison of the mean levels of input and output will show.

Clamping to any Level

The examples considered so far clamp the output waveform either positively or negatively to *zero volts*. In some radar circuits the output must be clamped to a level *other than zero*. The level to which the peak or the base of the waveform is clamped is determined by the voltage to which R and the appropriate diode electrode are returned. In every case so far R has been connected to earth (zero volts). Let us now consider some examples of clamping to other levels.

A circuit which gives *negative* clamping to a *positive* bias level is illustrated in Fig. 7a. Because of the inclusion of the bias voltage V_B two points should be noted:—

- $V_{in} = V_C + V_R + V_B$ at all times.
- $V_{out} = V_R + V_B$ at all times.

In Fig. 7b before the input is applied, C charges to the bias voltage V_B (in this example $-25V$) and remains steady at this value. V_R is zero and $V_{out} = +V_B = +25V$. Now let us consider the action from this point:

A to B. The input rises by $100V$ from zero. Since C cannot charge immediately V_C remains at $-25V$ and the voltage step of $100V$ is developed across R. Thus V_R rises *immediately* to $+100V$, and the diode cuts on. The output V_{out} rises to $100 + 25 = 125V$.

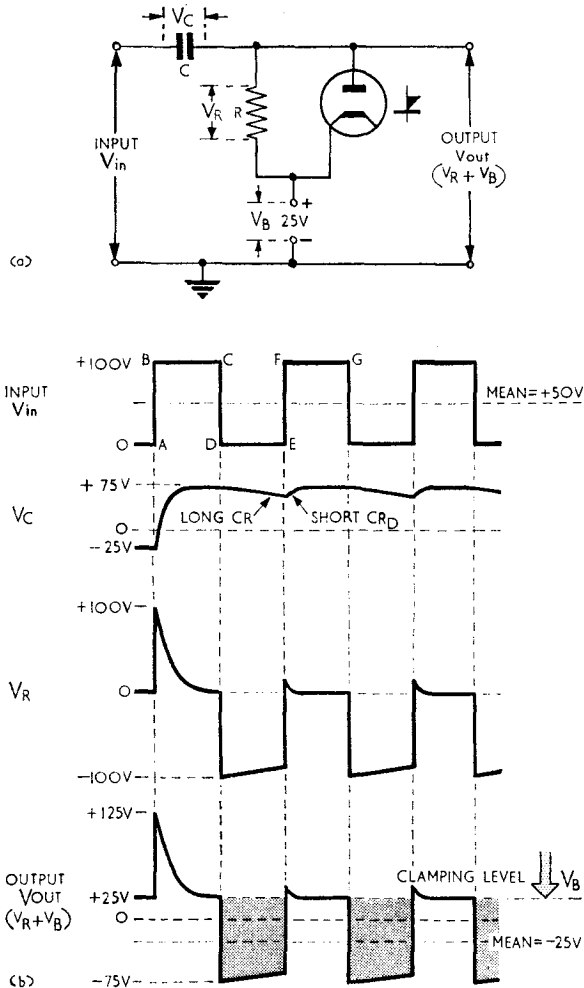


FIG 7. NEGATIVE CLAMPING TO A POSITIVE BIAS, CIRCUIT AND WAVEFORMS

B to C. With the diode conducting, C charges on a *short* time constant of CR_D seconds and V_C rises *quickly* to $+75V$ as V_R falls to zero. When V_R reaches zero the output voltage V_{out} has fallen to $+25V$. The diode is now *cut off*.

C to D. The input falls by $100V$ to zero. Since V_C cannot change instantly both V_R and V_{out} fall by $100V$, V_R to $-100V$ and V_{out} to $-75V$. The diode remains cut off and V_C remains at $+75V$.

D to E. Since the diode is cut off, C can only discharge very slowly on a *long* time constant of CR seconds and V_C falls *slightly* from $+75V$. Both V_R and V_{out} *rise* by the same amount as V_C falls.

E to F. The input rises by $100V$ from zero. V_R therefore rises by the same amount and so does V_{out} , V_C remaining constant. V_R rises to a slightly *positive* value and the diode *cuts on*.

F to G. With the diode conducting, C recharges on a *short* time constant of CR_D seconds and V_C rises *quickly* to $+75V$. V_R returns to zero in the same time (producing the usual pip) and V_{out} returns to the bias level of $+25V$.

Thereafter the cycle is repeated and the output is *negatively* clamped to the *positive* bias level of $+25V$. Examination of Fig. 7b will show that the output waveform is merely that of V_R superimposed on the bias level.

If the bias polarity is *reversed* then the output waveform is *negatively* clamped to a *negative* bias level of $25V$ (Fig. 8a). *Positive* clamping to a bias voltage is achieved simply by reversing the

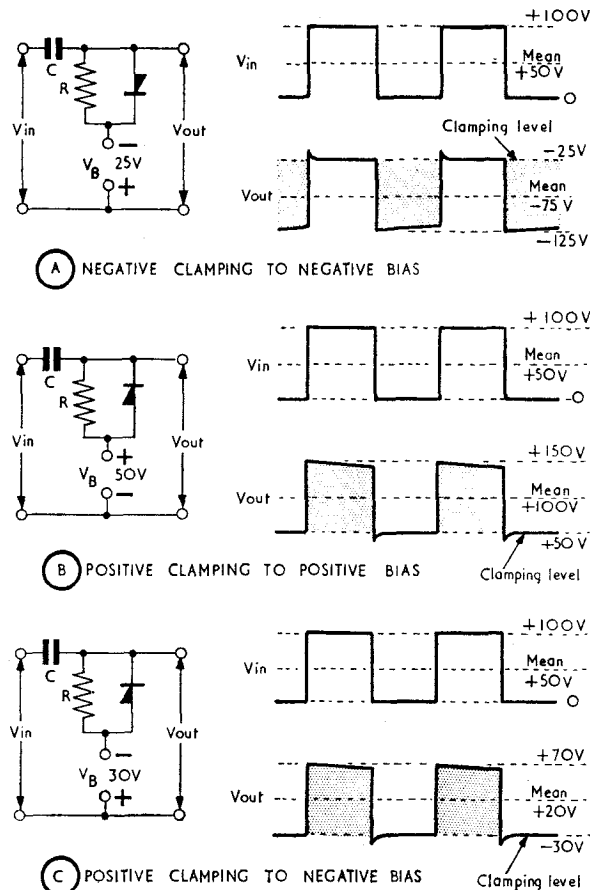


FIG. 8. POSITIVE AND NEGATIVE CLAMPING TO ANY LEVEL

connections to the diode. Fig. 8b shows *positive* clamping to a *positive* bias level of $50V$ and Fig. 8c shows *positive* clamping to a *negative* bias level of $30V$.

In practice, positive or negative clamping to almost *any* voltage level is required. We shall see examples of this later in these notes.

Summary of Diode Clamping Circuits

A waveform is *positively* clamped if it is *positive-going* from the clamping level. It is *negatively* clamped if *negative-going* from the clamping level. A simple rule for deciding quickly if a circuit is a positive or a negative clamp is to imagine the anode of the diode as a *piston* about to begin a stroke towards the cathode, pushing the waveform in the process. Then if the piston stroke is 'down' towards the cathode the waveform is being pushed *downwards* and the circuit is a *negative* clamp (Fig. 9a). If the movement is *upward*, as in Fig. 9b., the circuit is a *positive* clamp. The clamping level in each case is the voltage to which the resistor R is returned.

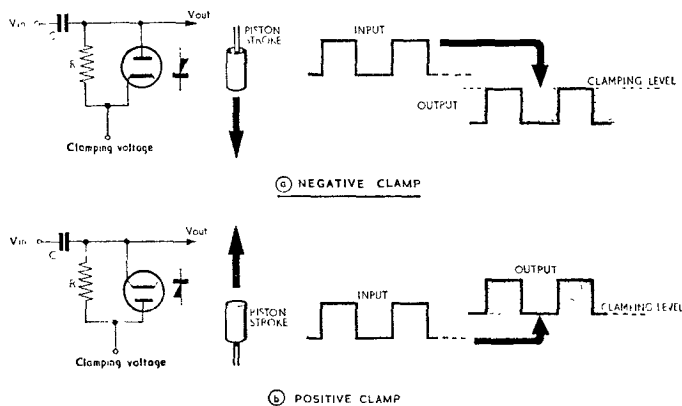


FIG 9. THE PISTON RULE FOR CLAMPING CIRCUITS

Although all the square wave inputs considered have been *symmetrical* the clamping circuits will work equally well with *asymmetrical* inputs provided the short CR circuit formed by the conducting diode and C has a time constant less than one-tenth of the pulse duration of the narrow pulse.

In addition, note that *sine wave* inputs may be clamped to a required voltage level in exactly the same way as square wave inputs.

Grid Current Clamping

Clamping may occur in the grid circuit of a triode or pentode amplifier if grid current is allowed to flow, because the grid-cathode path acts as a diode in the same way as in the grid current limiter (see p. 88). This clamping at the grid may be wanted or it may be unwanted.

If a waveform is applied through a long CR circuit to the grid of an amplifier operating at zero bias (Fig. 10a), grid current flows and the coupling capacitor is charged in the same way as that of the capacitor in a diode clamping circuit. Thus, with no bias, the positive peak of an incoming signal is clamped to zero volts and the input waveform is said to be *negatively clamped to zero volts* (Fig. 10b). The action is very similar to that for obtaining automatic grid leak bias in an oscillator.

Note that grid current clamping does not occur in a normal Class A amplifier because grid current is not normally allowed to flow. Even if grid current did flow to cause clamping at the grid the amplified output is taken from a different part of the circuit so that it has its own d.c. level. It is only the *input* waveform which may be clamped to zero volts.

If the cathode follower amplifier shown in Fig. 10c is used, *positive* clamping to a variable bias level $+V$ may be obtained. The negative peak of an incoming signal is now clamped to the voltage V .

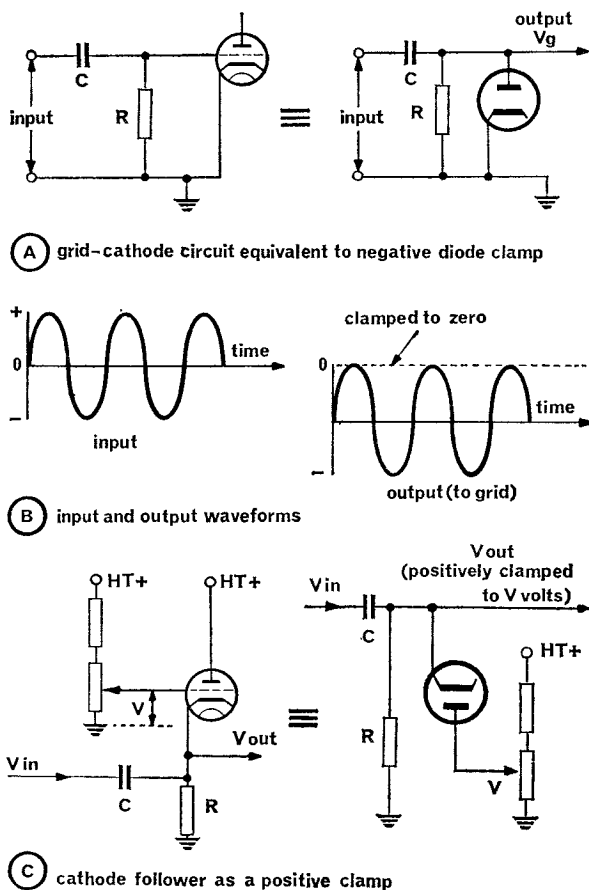


FIG 10. GRID CURRENT CLAMPING, CIRCUIT AND WAVEFORMS

Applications of Clamping Circuits

Clamping circuits have many uses in radar. The following paragraphs consider a few examples.

a. Pulse shaping. In Chapter 4 (p 88) we saw that the output from a triode squarer is not a perfect square wave. It can however be made more square by other circuits. We often need to 'tidy up' a waveform in this way and the process is known as *pulse shaping*, and may include both limiting and clamping. Fig 11 shows an example. The output from the squarer varies between $+200V$ and $+300V$ and is considerably curved below the $+225V$ level. This waveform is *positively clamped* to $-25V$ so that the curved portion is negative with respect to earth (zero volts) and the remainder is positive. The clamped waveform is then *negatively limited* to zero volts so that the curved portion is cut off and we are left with a good square wave of $75V$ amplitude, positive-going with respect to earth.

b. CRT blanking pulse. A blanking pulse is normally applied to the grid of a c.r.t. to cut off the electron beam during the resting time between traces. The required input V_g to the grid of the c.r.t. is shown in Fig 12b, where V_T is the level of voltage needed during the trace to give the required brightness. The voltage V_T is obtained from a voltage divider

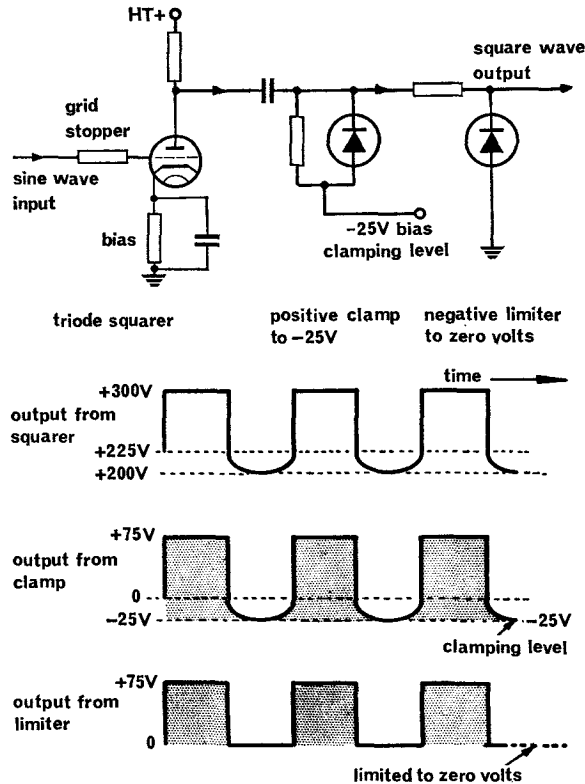


FIG 11. PULSE SHAPING, CIRCUIT AND WAVEFORMS

network, and the square wave input for the blanking pulse is applied to the grid through the usual CR coupling network. However only a.c. voltages appear across R and the square wave input will vary V_g above and below the bias level V_T by equal amounts (Fig 12c). The brightness of the trace will therefore increase and the electron beam *may not be blanked out* between traces. This may be avoided by using a *negative clamping circuit* to clamp the input to the bias level V_T (Fig 12d). The use of a clamping circuit has another advantage. If the *symmetry* of the square wave input changes, e.g. due to a change in the range scale, the *mean value* of the input will tend to vary causing a variation in the brightness of the trace. Clamping prevents this.

c. Deflection modulation. In an electrostatic c.r.t. the vertical position of the trace in a type A display, in the absence of a signal input, is determined by the shift voltage applied to the Y deflecting plates of the c.r.t. If only a few signals are being received the *mean level* of the voltage applied to the Y plates increases *slightly* (Fig 13a). If *many* signals are received the level *rises still further*. The result is that the trace tends to 'jitter' up and down as the *number* of signals being received varies. This can be prevented by *negatively clamping* the input waveform to the shift voltage (Fig 13b). Fig 13c shows the basic circuit.

d. Intensity modulation. This is the type of modulation used in a p.p.i. display. In this case, in the absence of clamping, variation in the *number* of signals applied to the c.r.t. grid varies the overall *brightness* of the display because of the variation in the mean level of the input (Fig 14a). This difficulty is overcome by *positively clamping* the minimum signal

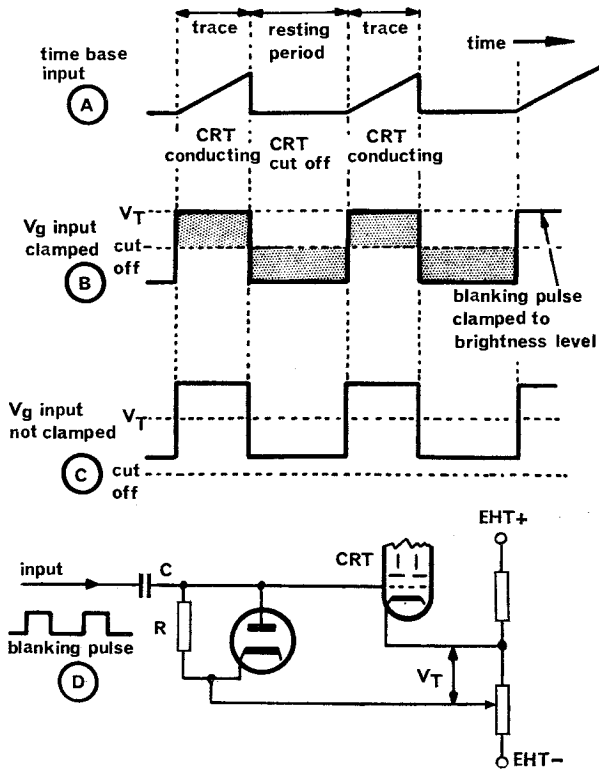


FIG 12. CLAMPING OF CRT BLANKING PULSE

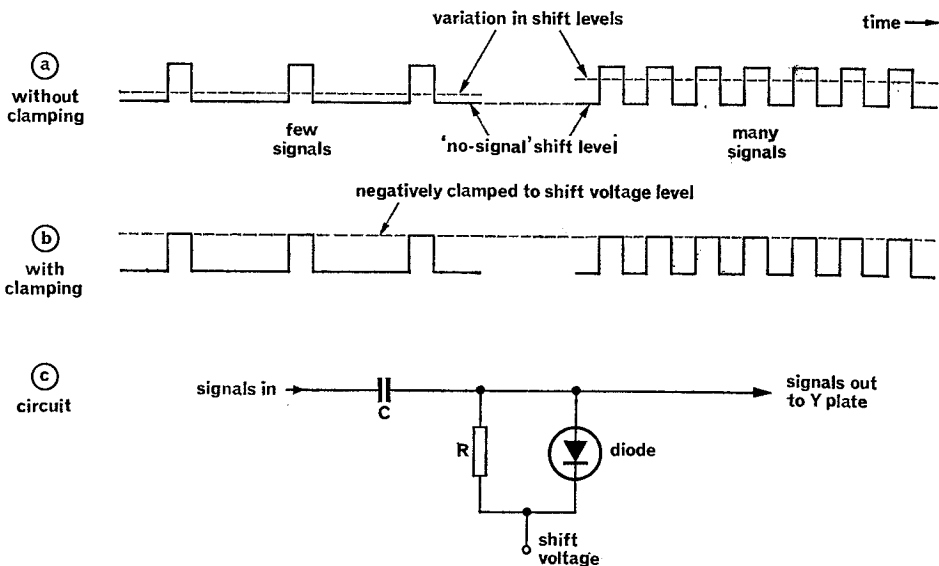


FIG 13. CLAMPING IN DEFLECTION-MODULATED DISPLAYS

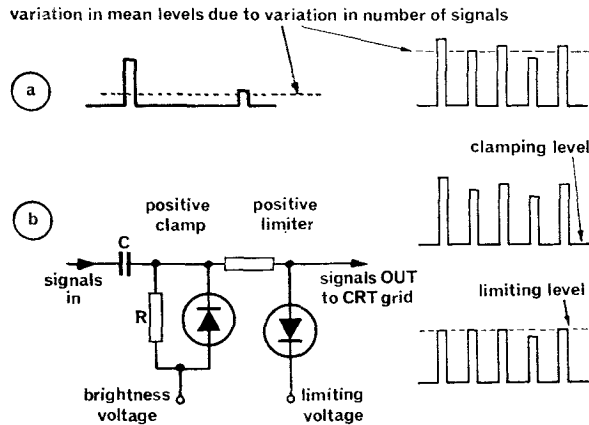


FIG 14. CLAMPING IN INTENSITY-MODULATED DISPLAYS

voltage to a fixed value (set by the brightness control) as shown in Fig 14b. In addition a positive limiter is included in the circuit. This ensures that all signals produce the same brightness of echo whatever their strength. If signals were not limited in this way the weak echoes would be masked by the stronger ones.

CHAPTER 6

FREE-RUNNING (ASTABLE) MULTIVIBRATORS

Introduction

The importance of square waves for timing and for controlling the action of the circuits in a radar installation has already been mentioned. So far we have learnt the terms used to describe square waves and have seen the effects of applying square waves to various types of circuit. The next step is to discover how square waves are *produced* (Fig 1).

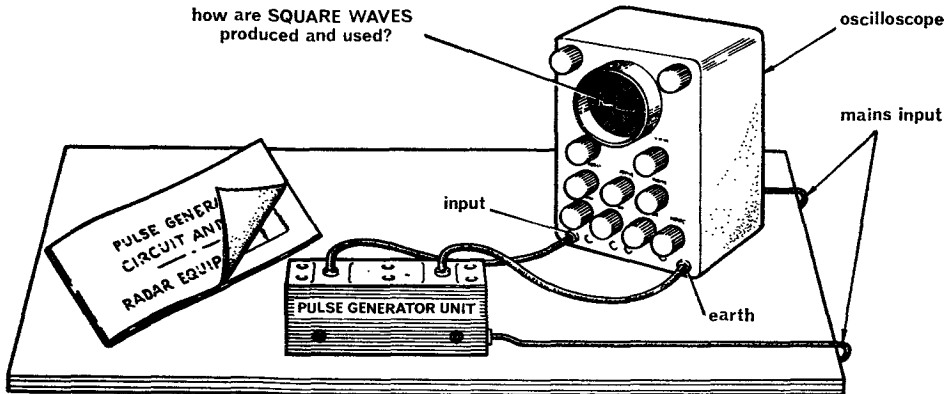


FIG 1. PRODUCTION OF SQUARE WAVES

In Chapter 4 we saw how limiting circuits may be used to *convert* sine wave inputs to approximate square wave outputs. Apart from the fact that this is a rather indirect way of obtaining a square wave, the output waveform is not sufficiently steep-sided for accurate timing and precise triggering. If the output from a squarer is being used to trigger a succeeding stage, any variation in the bias of that stage will alter the instant of triggering (Fig 2a). With a steep-sided square wave input this does not happen (Fig 2b).

The circuits used to produce steep-sided square wave outputs are known as *square wave generators* or *pulse generators*. Most of these circuits have two valves (or two transistors) which are coupled by CR combinations in such a way that the valves switch each other on and off at regular intervals, giving square waves of voltage at the output. The term "square waves" is being used here very loosely. The output may in fact be *symmetrical* or *asymmetrical*. Many different circuits are in common use as square wave generators and some of these are discussed in this chapter.

Some Important Basic Facts

Most pulse generators depend for their operation on a combination of several simple basic facts, all of which we already know. The action of some of these

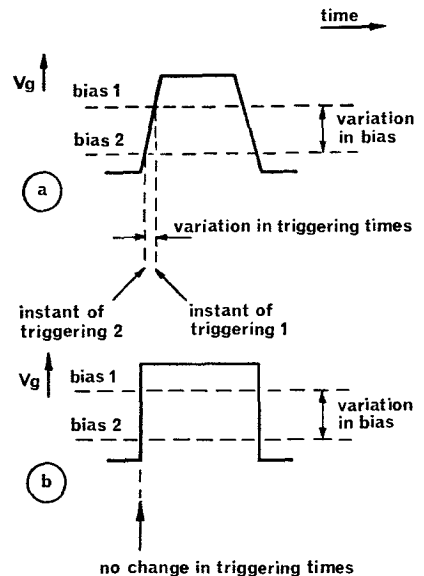


FIG 2. NEED FOR STEEP-SIDED SQUARE WAVES

circuits may appear, at first, quite complicated. They appear so only because several actions may be taking place simultaneously, or with very short time intervals between them. Each action considered by itself is basically simple and if we consider each step in isolation and in the correct sequence, the action of the whole becomes quite clear. A few of the basic actions which occur time and again in pulse circuits are summarized below:

- If the control grid of a pentode valve is below cut-off, no valve current flows whatever the voltages on the screen, suppressor or anode.
- If the control grid rises above zero volts with respect to cathode, grid current flows. The grid-cathode resistance is then very low, typically $1k\Omega$. If a high resistance is connected between the grid and a point of high voltage, grid limiting takes place to 'clamp' the grid to just above zero volts.
- If a valve is conducting, its anode voltage V_a is decided by its anode current I_a , the load R_L and the value of h.t. Thus if I_a is $5mA$, R_L $10k\Omega$ and h.t. $300V$, then V_a must be $250V$ whatever other parts of the circuit may be doing. If a valve is cut off, its anode voltage rises, usually to the value of the h.t. supply; it may rise even *above* h.t. if 'ringing' takes place (see p 81).
- CR series networks form an important part of most pulse generator circuits. The effects of applying sudden changes of voltage to such circuits are discussed in detail in Chapter 2 of this section (p 67). There we said that a capacitor cannot change its charge, and hence the

voltage across it, instantaneously. Thus if $100V$ is suddenly applied to C and R in series the full $100V$ appears across R initially; C then commences to charge on a time constant CR seconds. These statements also apply if there is more than one resistor connected in any series order with C ; the $100V$ is initially shared between the resistors according to their value. If the capacitor is initially charged to, say, $20V$ before the step is applied then the *difference* ($80V$) is initially shared between the resistors (Fig 3b). In each case as C charges so the voltage across it increases and the voltages across the resistors decrease (Fig 3c).

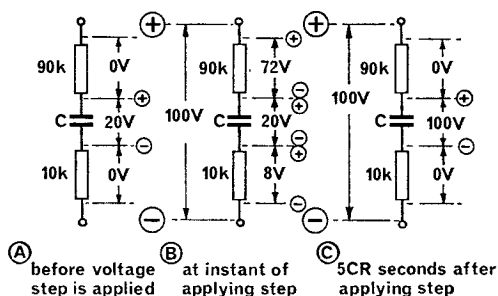


FIG 3. CR SERIES CIRCUITS

- The combination of valves and CR circuits forms the basis of most pulse generators. Suppose a CR network is connected to two valves as shown in Fig 4. If V_1 is conducting, its V_a is decided by the valve current, by R_{L1} , and by the ht. If V_a is $+30V$ ($170V$ developed across R_{L1}) and remains steady at this value then C_1 will charge to $30V$. There is then no voltage across R_{g1} , and point B is at zero volts (Fig 4a). If now V_1 is suddenly cut off we have $170V$ (the ht of $200V$ less the $30V$ across C_1) shared between R_{L1} and R_{g1} , in proportion to their values. Thus, at the instant V_1 cuts off point A rises to $+185V$ ($15V$ dropped across R_{L1}) and point B rises to $+155V$ ($155V$ across R_{g1}). C_1 is unaffected instantaneously and remains charged to $30V$ (Fig 4b). C_1 now commences to charge through R_{L1} and R_{g1} and will eventually charge to $200V$; the voltages across R_{L1} and R_{g1} will then both be zero. Point B is thus at zero volts and A at $+200V$ (Fig 4c).

In this simplified explanation we have ignored the effect that V_2 has on the circuit when this valve is suddenly switched on. We shall consider this point, and others, later.

However, the important factors to bear in mind are that when one valve is suddenly cut off its anode voltage rises rapidly towards ht+. This full rise of voltage is transferred to the grid of the *other* valve through the capacitor (which cannot change its charge instantaneously) and this

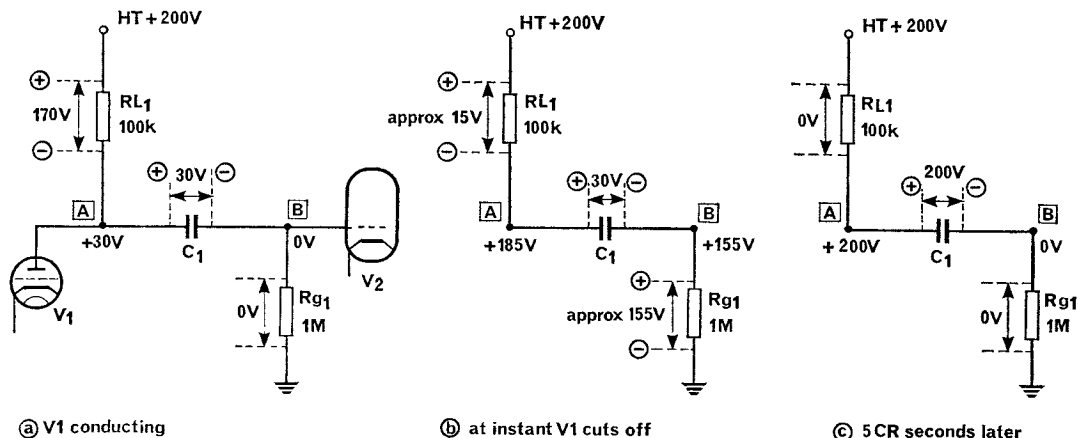


FIG 4. CR NETWORK CONNECTED TO VALVES

other valve is immediately driven hard on. We thus have switching between two valves. Somewhat similar remarks apply also to transistors.

Multivibrators

Multivibrators are the most commonly used square wave generators and they belong to a family of oscillators called *relaxation oscillators*. A multivibrator has two different electrical states, usually high voltage output and low voltage output, and there are three main types of such *two-state* circuits:

a. **Astable multivibrator.** This switches continuously between the two states at a constant repetition rate *without external triggering*. It is therefore a 'free-running' oscillator and it produces continuous square waves at its output.

b. **Monostable multivibrator.** This circuit has only *one* stable state. When it is *triggered* by an external signal it passes into the other state, remains there for a given time and then returns, *of its own accord*, to the first (stable) state. It will remain in the stable state until another triggering input pulse is received. This circuit is also referred to as a "one-shot multivibrator" because it goes through a full cycle in response to a *single* triggering pulse. A more common name however, and the one we shall use in these notes, is "flip-flop".

c. **Bistable multivibrator.** In this circuit *both* states are stable, the circuit remaining in *either* of them indefinitely. It passes from one state to the other only when suitably triggered, and to go through a full cycle *two* triggering pulses are needed. This circuit is also known as a bistable trigger circuit, a toggle or an Eccles-Jordan circuit.

Note that only the astable multivibrator is free-running. The others must be triggered.

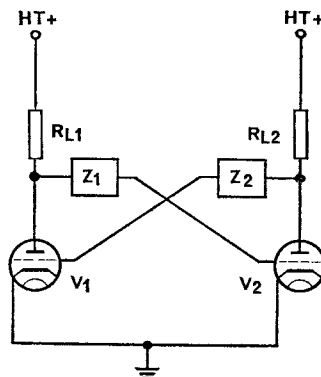


FIG 5. BASIC MULTIVIBRATOR CONNECTIONS

The basic circuit of a multivibrator using triodes is illustrated in Fig 5. The important point to notice is the *cross-coupling* between the anode of one stage and the grid of the other. This form of coupling ensures that the input of each stage is derived from the output of the other. Thus the output from V_1 is applied through an impedance Z_1 to the grid of V_2 . V_2 provides an amplified and *phase-inverted* output which is fed back through an impedance Z_2 to the grid of V_1 . The signal is again amplified and *phase-inverted* through V_1 so that an initial signal at V_1 anode reappears at the same point greatly amplified and in the *same phase*. The circuit therefore tends to oscillate. Whether it does so or not depends upon the *nature* of the cross-coupling impedances Z_1 and Z_2 . In general, if they are both *capacitances* the circuit is that of a free-running *astable* multivibrator. If one is a *capacitance* and the other a *resistance* the circuit is *monostable*. If both are *resistances* the circuit is *bistable*. We shall now consider the first of these circuits in more detail. The other two circuits, which require a triggering input, are dealt with in the next chapter.

Free-running Anode-coupled Multivibrator

Fig 6 shows the circuit diagram of an astable multivibrator and also the waveforms which appear at the anodes. The action of the circuit, as of all relaxation oscillators, can be divided into two phases:

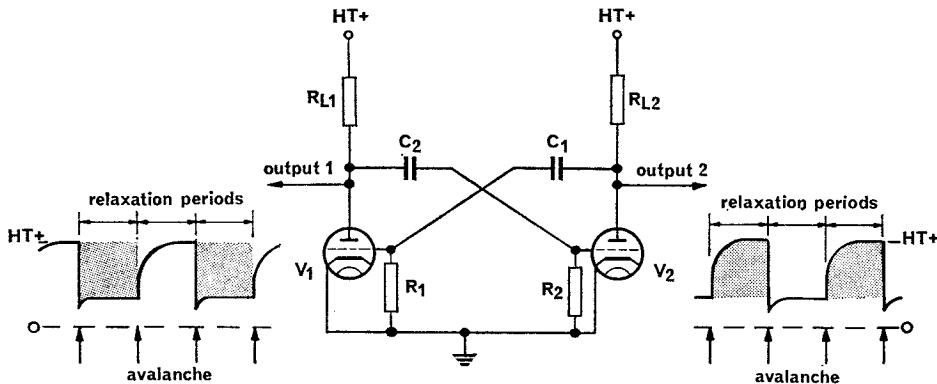


FIG 6. BASIC ASTABLE MULTIVIBRATOR

- a. The periods corresponding to the vertical parts of the waveforms, when both valves are conducting and acting as amplifiers with positive feedback from each anode to the other grid. This condition is unstable and initiates a *cumulative action* or *avalanche* which very quickly drives one valve well beyond cut-off and causes the other to conduct heavily.
- b. The periods corresponding to the 'flat' parts of the waveforms during which one valve is conducting heavily and the other is cut off but approaching the point where it will cut on. This is the *relaxation period*.

Let us imagine that the circuit is approaching the end of a relaxation period and that V_2 is conducting heavily (with its grid clamped to zero volts by grid current limiting) and V_1 is cut off but rising towards its cut-off point. With V_2 conducting, its anode voltage will be at a low working value, depending upon the valve current and the values of R_{L2} and h.t. Also, since V_1 is still cut off, V_1 anode will be at the same voltage as h.t. + (no current through R_{L1} and no voltage drop across it). This condition before the start of an avalanche, is illustrated in Fig 7a.

When the voltage at V_1 grid rises above cut-off, V_1 anode voltage *falls* by an amount depending upon the gain of the stage . . .

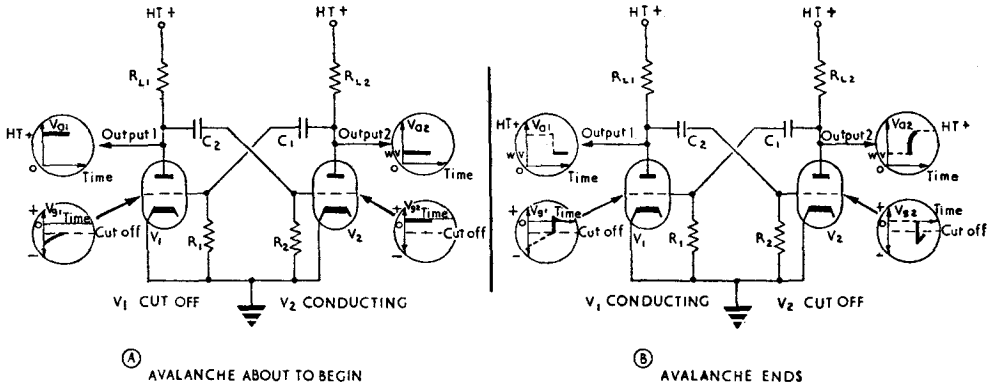


FIG 7. AVALANCHE EFFECT IN MULTIVIBRATOR

... This fall is transferred through C_2 to V_2 grid and, since C_2 cannot change its charge instantly V_2 grid voltage *falls* from zero to a *negative* value causing V_2 anode voltage to *rise* ...

... This rise is transferred through C_1 to V_1 grid causing V_1 grid to rise further above cut-off and V_1 anode voltage to *fall even more* ...

... This further fall in V_1 anode voltage is again transferred through C_2 to V_2 grid and the action is *cumulative*.

The result is that the voltages at V_1 grid and V_2 anode *rise* very rapidly and those at V_2 grid and V_1 anode *fall* very rapidly, giving the avalanche effect. Because of the rapidity of the action and the cumulative amplification that takes place, the time for such an avalanche is extremely small—a fraction of a microsecond—and steep-sided waveforms are produced. At the end of the avalanche the roles of the two stages are reversed, i.e. V_1 is now conducting and V_2 is cut off (see Fig 7b).

During the relaxation period which follows the avalanche two things happen simultaneously:

a. With V_2 cut off, V_2 anode voltage tries to rise to h.t. + but it can only do so *exponentially* as C_1 charges through R_{L2} (see Fig 8a). This gives the curve in V_2 anode waveform.

b. V_1 is conducting and its anode voltage is now steady at its working value. C_2 was initially charged to the h.t. voltage and now starts to *discharge* through R_2 (Fig 8b). V_2 grid voltage therefore rises *exponentially* from its negative value towards earth (zero volts) with a time constant of $C_2 R_2$ seconds. When V_2 grid voltage rises above cut-off, V_2 is then

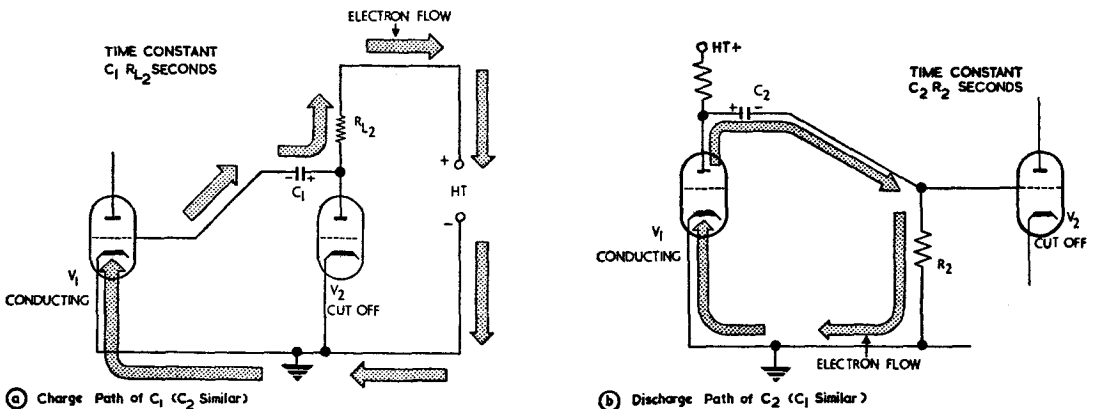


FIG 8. CAPACITOR CHARGE AND DISCHARGE PATHS

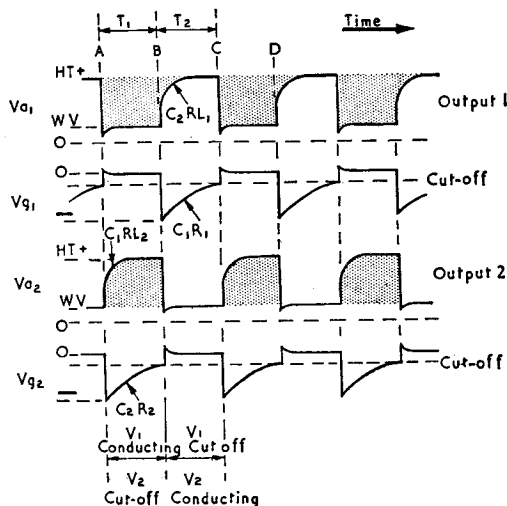


FIG 9. BASIC MULTIVIBRATOR WAVEFORMS

in exactly the same state as was V_1 at the beginning of the action. Another avalanche then takes place and the action is repeated continuously.

The series of events may be summed up with reference to the waveforms of Fig 9:

Instant A. V_{g1} rises above cut-off and V_{a1} falls ...

... V_{g2} falls (through C_2) and V_{a2} rises ...

... V_{g1} rises further (via C_1) and an avalanche results. V_{a1} falls to its working value and V_{g2} falls by the same amount. V_{g1} rises to just above zero, where it is held by grid current limiting, and V_{a2} rises by the same amount.

Interval A to B. In the relaxation period C_2 discharges through R_2 and V_{g2} rises exponentially towards zero; C_1 charges through R_{L2} and V_{a2} rises exponentially to h.t.+. V_{g1} is held at zero volts by grid current limiting and V_{a1} is steady at its working value.

Instant B. V_{g2} rises to cut-off and the action is the same as that described for instant A with the roles of the valves reversed. Thereafter the action is repeated.

From each anode we have anti-phase 'square wave' voltages. The shapes of these waveforms may be improved by applying the outputs to limiting and clamping circuits. The 'pips' in the V_g waveforms, which are reflected also in the V_a waveforms, are caused by grid current limiting which 'pulls' the grids back to zero volts. They may be reduced by inserting grid stoppers in each stage.

Control of Multivibrator Frequency

The time period T_1 in Fig 9, during which V_2 is cut off, corresponds to the time taken for C_2 to discharge through R_2 to the cut-off value and is determined by the time constant C_2R_2 seconds. Similarly, T_2 is determined by C_1R_1 .

If both stages are using identical components then $C_1R_1 = C_2R_2$ and $T_1 = T_2$. The waveform of voltage at either anode is then *symmetrical*. If $T_1 = T_2 = 1000 \mu s$, the time taken for one complete cycle is $2000 \mu s$ and the frequency of the output is $\frac{1}{2000 \mu s}$ or 500 c/s. Thus the frequency at which a free-running astable multivibrator oscillates is determined by the time constants C_1R_1 and C_2R_2 and depends upon the relationship:

$$\frac{1}{(C_1R_1 + C_2R_2)}$$

To change the frequency, any of the four components may be altered. However, if we wish to change the frequency *without affecting the symmetry* we must change *both* time constants *by the same amount*. This may be done by the frequency control R_f shown in Fig 10.

Control of Output Symmetry

If one valve in a multivibrator remains cut off for a longer period than the other, T_1 no longer equals T_2 and the output becomes *asymmetrical*. An asymmetrical output can be obtained fairly easily by making the time constant C_1R_1 different to that of C_2R_2 . However, in practice we often need to be able to vary the symmetry *without affecting the frequency*. This may be done

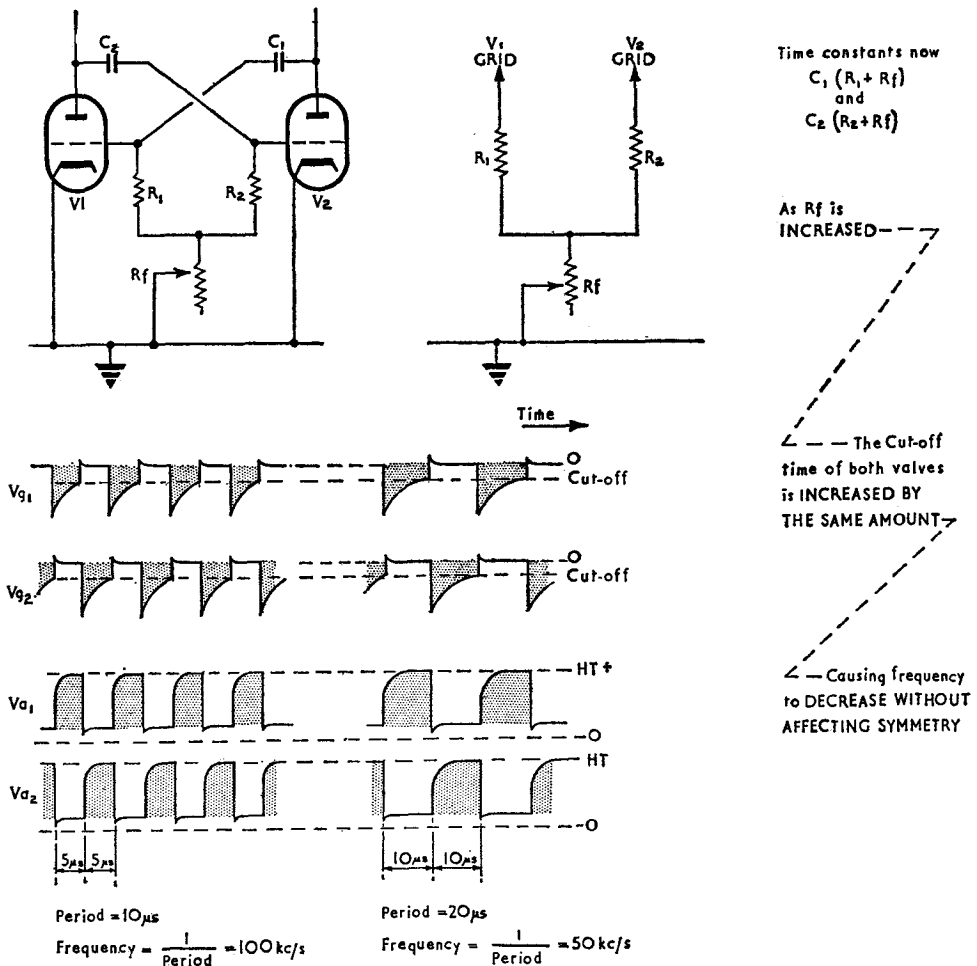


FIG 10. CONTROL OF MULTIVIBRATOR FREQUENCY

by the symmetry or pulse duration control R_s shown in Fig 11. When the wiper arm of R_s is at the centre position, the resistance of both CR circuits is the same and the output is symmetrical. The effect of moving the wiper from left to right is illustrated in Fig 11. As the time constant $C_2 R_2$ decreases, that of $C_1 R_1$ increases by an equal amount. Symmetry is therefore varied without affecting the frequency.

Modifications to the Basic Circuit

In a multivibrator the change-over from the relaxation period to the avalanche occurs when the grid voltage of the non-conducting valve rises above cut-off. In the circuits considered so far we have shown the grid leaks R_1 and R_2 connected to earth. The coupling capacitor connected to the grid of the cut-off valve is therefore discharging towards earth and is said to be 'aiming at zero volts'. Since the cut-off point is relatively near zero volts the capacitor discharge is nearing the 'flat' part of its exponential curve. This slow rise through cut-off gives an indeterminate cut-off point, because any slight variations in the circuit conditions (e.g. variation in h.t. supply) will cause the cut-off point to be reached sooner or later than anticipated. Frequency and symmetry may thus vary (see Fig 12).

This difficulty can be overcome by returning the grid resistors to h.t. + instead of to earth.

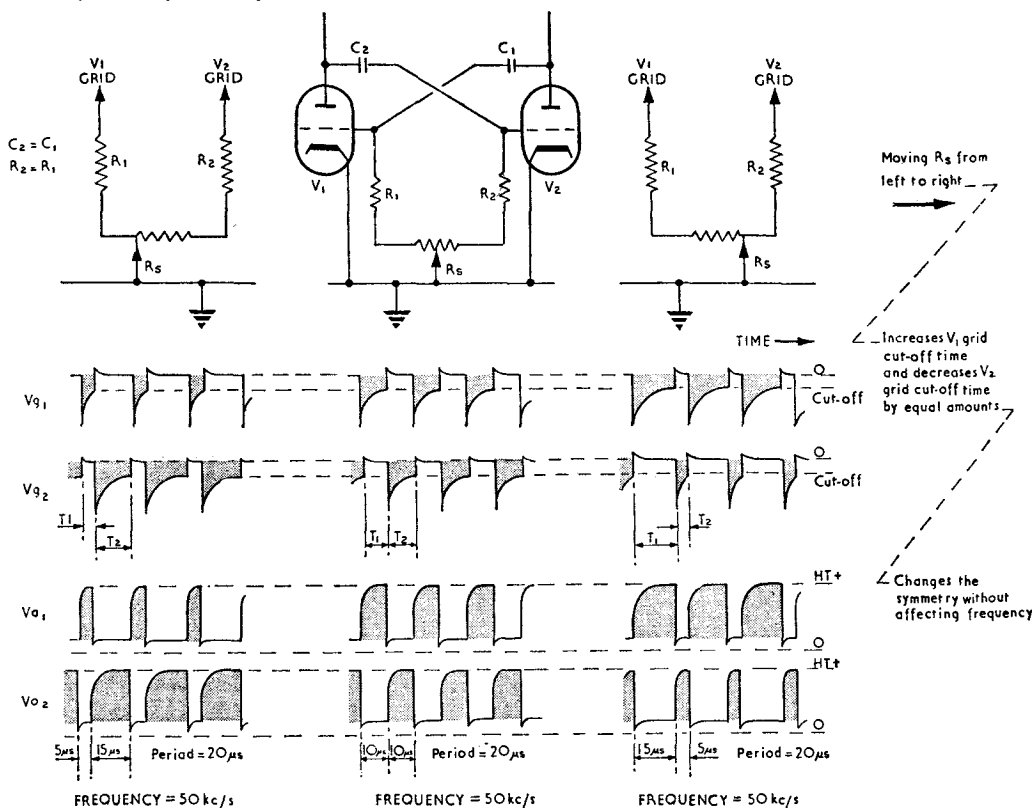
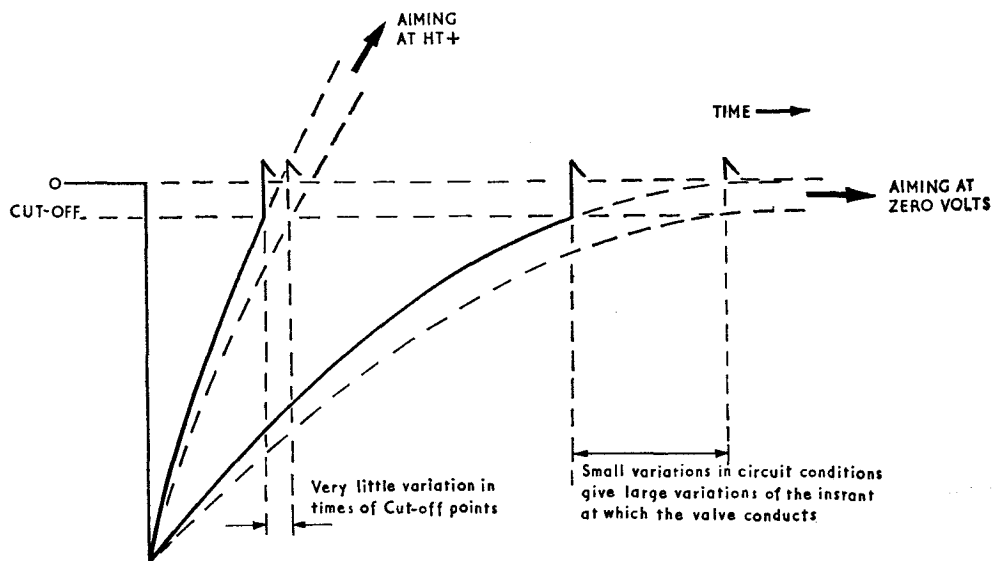


FIG 11. CONTROL OF MULTIVIBRATOR SYMMETRY



NOTE THAT VARIATION OF AIMING VOLTAGE WILL VARY FREQUENCY

FIG 12. EFFECT OF A POSITIVE AIMING VOLTAGE

The capacitors are now *aiming at h.t. +*. The result is that when the capacitor voltage rises through cut-off, the capacitor discharge is still at the 'steep' part of its exponential and this steep rise through cut-off gives a more decisive cut-off point which is less subject to random variations in the circuit conditions.

The circuit of a multivibrator using this modification is shown in Fig 13. The circuit contains symmetry and p.r.f. controls. The symmetry control is as previously described. The p.r.f. control varies the 'aiming voltage' of both grids by equal amounts causing both valves to reach cut-off earlier or later depending upon the sense of variation. The frequency is thus varied without affecting symmetry.

Synchronized Multivibrator

If we apply a positive-going synchronizing pulse to the grid of one valve in a multivibrator before V_g would normally reach its cut-off point we can make the valve conduct at the instant the sync pulse is applied. This is shown in Fig 14. The frequency of the sync pulses must be slightly greater than that of the free-running multivibrator and their amplitude must be sufficient

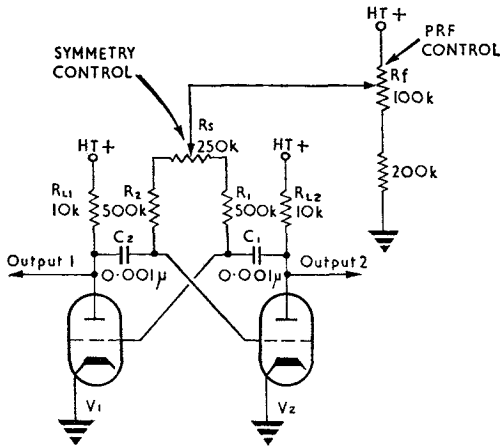


FIG 13. PRACTICAL MULTIVIBRATOR

SYNCHRONIZED PULSES TO V_1 GRID

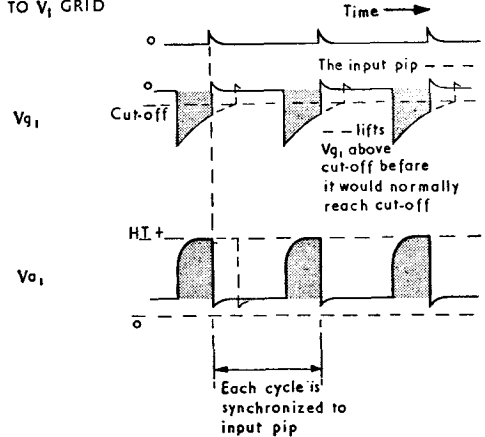


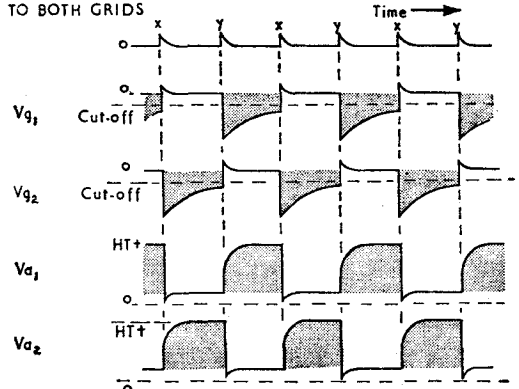
FIG 14. SYNCHRONIZED MULTIVIBRATOR

to take the grid above cut-off at the desired instant. Each cycle of the multivibrator is then 'locked' or synchronized to the frequency of the sync pulses. This method is often used to ensure that various circuits in a radar system all operate at the same instant of time. Note however that the astable multivibrator is *not dependent* for its operation on the sync pulses. It will still free-run at its own frequency without them, unlike the monostable and bistable multivibrators which must be triggered.

If we apply sync pulses to *both* valves, as in Fig 15, each *half-cycle* of the multivibrator output is synchronized. The sync pulse 'x' cuts on V_1 and pulse 'y' cuts on V_2 .

In some circuits the sync pulse is applied to the *anode* instead of directly to the grid. This

SYNCHRONIZED PULSES TO BOTH GRIDS



Pulse x lifts V_{g1} above cut-off before its normal time
Pulse y does the same to V_{g2}

BOTH HALF-CYCLES SYNCHRONIZED
FIG 15. SYNCHRONIZING EACH HALF-CYCLE

has the same effect, because the pulse is transferred from the anode to the appropriate grid through the coupling capacitor, which cannot change its charge instantaneously.

Electron-coupled Multivibrator

The output waveforms from a triode multivibrator have curved rising edges, due to the charging of the coupling capacitors, and have 'tails' on the falling edges, due to grid current limiting. The waveforms can be made more square by using pentodes arranged as an electron-coupled oscillator (Fig 16a). In this circuit the screen grids are acting as the 'anodes' of the

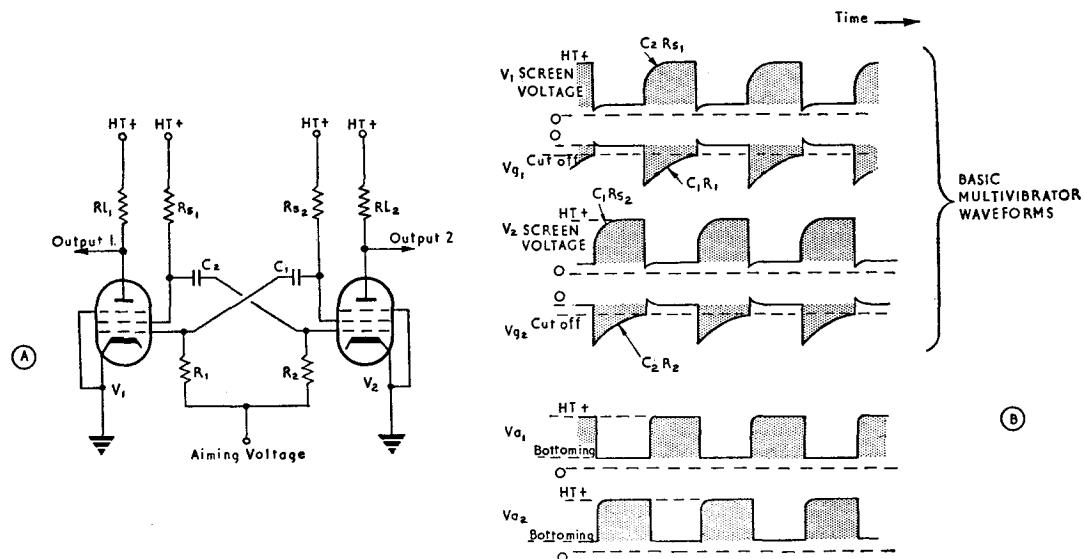


FIG 16. ELECTRON-COUPLED MULTIVIBRATOR

oscillator and the waveforms at the grids and screens are the same as those that appear at the grids and anodes of the triode multivibrator (see Fig 16b). As the grid voltage of a valve *both* rise towards h.t.+. However, there are no coupling capacitors connected to the anodes so that the *anode* voltages can rise almost *immediately*, whereas the screen voltages can only rise *exponentially* as their coupling capacitors charge. The outputs from the anodes are therefore much more square.

In addition, if the anode loads R_{L1} and R_{L2} are large enough to cause *bottoming* before V_g rises to zero volts, the positive pips on the grid waveforms are *not reproduced* as tails in the output waveforms.

The other advantage of the electron-coupled multivibrator is that the output is taken from an electrode which is not concerned in the multivibrator action and the load does not interfere with the operation of the circuit. Thus, variation in the load conditions has little effect on the multivibrator frequency, and stability is greatly improved.

Cathode-coupled Multivibrator

The circuit and waveforms of a cathode-coupled multivibrator are illustrated in Fig 17. In this circuit V_2 is connected to V_1 , not by the usual RC coupling from anode to grid, but by *feedback* through a common cathode resistor R_K , as in the long-tailed pair discussed in Part 1 of these notes. V_1 is a low-current valve, and V_2 passes a high current.

Let us assume that the circuit is oscillating with V_1 conducting and V_2 cut off but with its grid voltage just about to reach cut-off (say $-5V$). Then initially $V_{a2} = \text{h.t. (+300V)}$; V_K is a

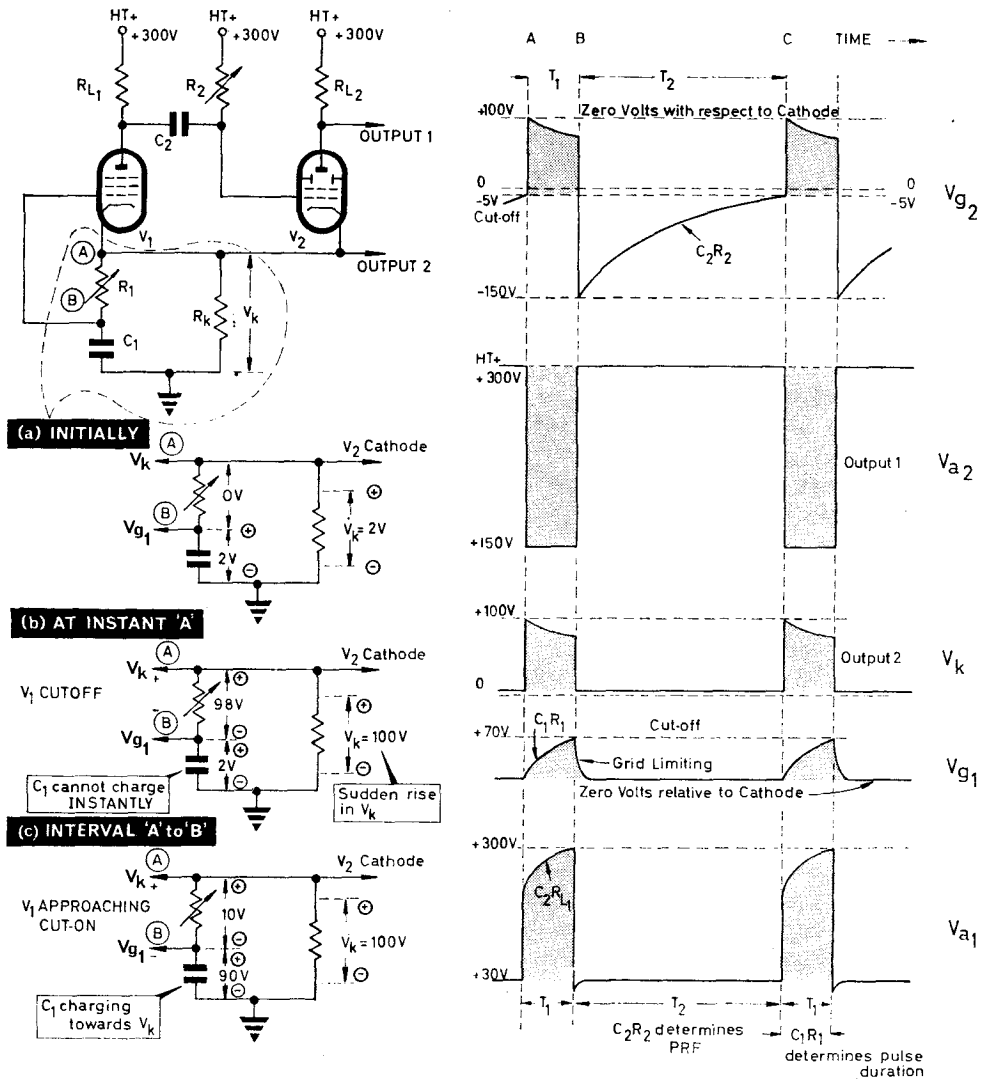


FIG 17. CATHODE-COUPLED MULTIVIBRATOR, CIRCUIT AND WAVEFORMS

low value (say + 2V) because of the small current passed by V_1 ; V_{g1} is held at zero volts with respect to its cathode, because there is no voltage drop across R_1 (see Fig. 17a); and V_{a1} is at its low working voltage (say + 30V).

Instant A. V_{g2} rises to cut-off . . .

. . . V_2 conducts causing V_{a2} to fall and V_K to rise (because of the increased current from V_2 through R_K) . . .

. . . V_{g1} is held constant at just above zero volts by C_1 , which cannot charge instantly. The sudden rise in V_K is therefore developed across R_1 (see Fig. 17b) and, with point A going positive to point B, V_1 cuts off . . .

. . . V_{a1} thus rises towards h.t. and this rise is transferred through C_2 to cause V_{g2} to rise also . . .

. . . V_2 conducts more heavily causing V_{a2} to fall further.

This cumulative action *cuts* off V_1 and leaves V_2 *conducting* heavily. The avalanche ceases when V_1 is cut off by the rise at point A, point B being held constant by C_1 . V_1 grid is then sufficiently *negative* with respect to its cathode to cut off V_1 . V_{a1} and V_{g2} have both risen by the same amount, say 100V—much larger than the initial rise in a simple multivibrator because of the cathode follower action of V_2 . V_K has also risen to about + 100V and V_{a2} has fallen to its working value, say + 150V.

Interval A to B. With V_1 cut off, C_2 *charges* through R_{L1} causing V_{a1} to rise exponentially towards h.t. + and V_{g2} to fall towards zero volts. As V_{g2} falls, V_K falls with it (due to the fall in current through R_K from V_2). The voltage across C_1R_1 is V_K and during this period C_1 *charges* through R_1 causing V_{g1} to rise towards V_K (see Fig. 17c).

Instant B. V_{g1} , rising towards V_K , reaches cut-off when the voltage across R_1 is sufficiently small . . .

- . . . V_1 conducts causing V_{a1} and V_{g2} (via C_2) to fall . . .
- . . . The fall in V_{g2} causes V_{a2} to rise and V_K to fall (less current through R_K) . . .
- . . . The fall in V_K (at point A) leaves V_{g1} (point B) *positive* with respect to cathode, since C_1 cannot discharge instantaneously . . .
- . . . V_1 thus conducts more heavily causing V_{a1} and V_{g2} to fall further.

A second avalanche now occurs which cuts off V_2 and leaves V_1 conducting. V_K falls immediately to a low value; V_{g1} is held *positive* by the charge on C_1 ; V_{a1} falls to its working value; and V_{g2} falls by the same amount well below cut-off; with V_2 cut off V_{a2} rises to h.t.+.

Interval B to C. C_1 is now *discharged* rapidly by the flow of grid current in V_1 and V_{g1} is clamped to zero volts with respect to cathode by this action. C_2 *discharges* through R_2 towards the aiming voltage and V_{g2} rises exponentially towards h.t.+.

Instant C. V_{g2} reaches cut-off and the action as from instant A is repeated.

A *negative-going* output may be taken from the anode of V_2 (V_{a2}) and a *positive-going* output from the common cathode (V_K). V_{a2} output has a good waveform because there is no capacitor connected to V_2 anode. The output from the cathode has a low source impedance due to the cathode follower action of V_2 and this may be useful for *matching*. Since *both edges* of the output waveforms are caused by an avalanche, very steep edges result—ideal for *timing*.

The period T_1 in Fig. 17 is determined by the time constant C_1R_1 and varying R_1 varies the *pulse duration*. The period T_2 depends upon C_2R_2 and varying R_2 (or the aiming voltage) varies the *p.r.f.* Since C_1R_1 can be made much smaller than C_2R_2 a very *asymmetrical* output may be obtained from this circuit.

Multivibrator using Transistors

In many radio circuits the thermionic valve can be replaced by its semiconductor equivalent, the transistor. A version of a basic multivibrator using transistors—a collector-coupled circuit—is shown in Fig. 18a.

As we have done throughout these notes we shall consider only the p-n-p transistor, although the n-p-n type could be used equally well if we reversed the polarity of the supply voltages. It will be remembered that the current through a p-n-p transistor *increases* as the base is made *more negative* relative to emitter. To cut off a p-n-p transistor we make the base *positive* with respect to emitter. Note that a valve cuts off when V_g falls below the cut-off point at a negative value equal to the grid base; a p-n-p transistor, on the other hand, cuts off *as soon as* its base goes positive with respect to its emitter, i.e. its 'grid base' is zero.

Because the polarities in a p-n-p transistor circuit are opposite to those in valve circuits, the waveforms associated with a p-n-p transistor are often drawn '*negative up*' as in Fig. 18b. This also means that the circuit action can be tied more closely with that of a circuit using valves. In some textbooks p-n-p transistor waveforms are drawn in the more conventional '*positive up*'

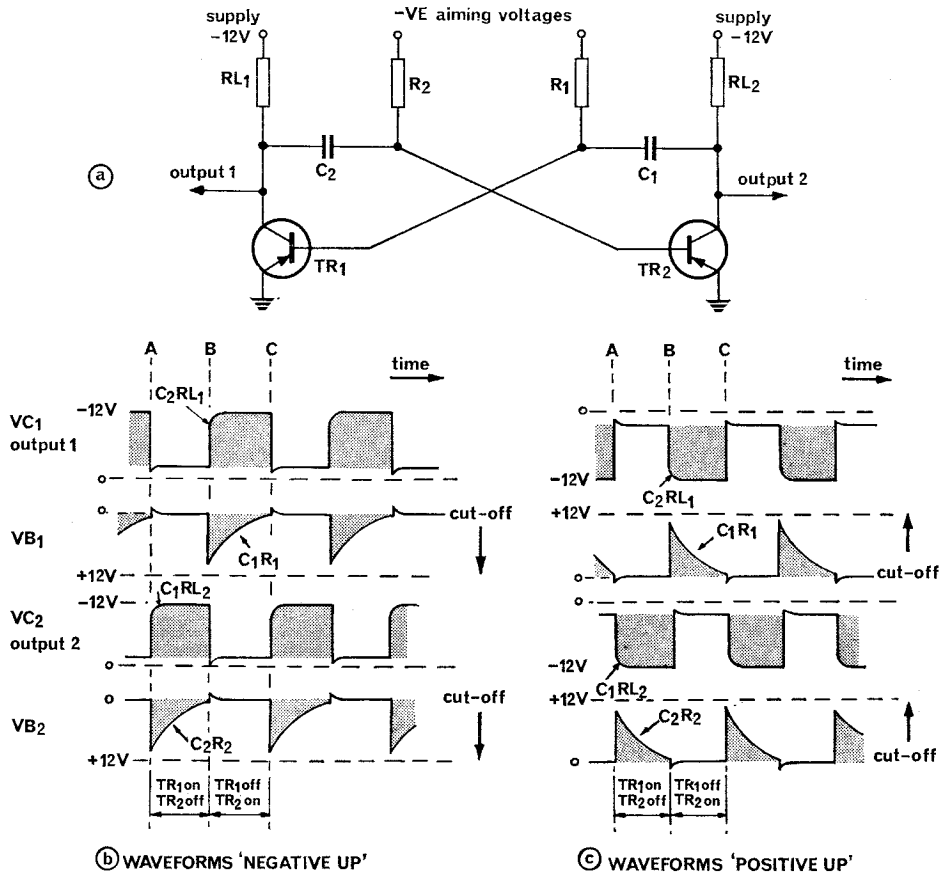


FIG 18. COLLECTOR-COUPLED MULTIVIBRATOR, CIRCUIT AND WAVEFORMS

style as in Fig 18c. Note however that the waveforms of Fig 18b and c are *exactly equivalent*. They are merely different ways of looking at the same thing. Throughout these notes we shall draw p-n-p transistor circuit waveforms 'negative up'. If for any reason 'positive up' versions are required, the waveforms are simply turned upside down.

The circuit action of a collector-coupled multivibrator is basically the same as that of the free-running astable multivibrator using valves. We shall assume that TR_2 is conducting and TR_1 cut off but with its base voltage just about to reach cut-off (zero volts). Then initially V_{C_2} is at its low working voltage, V_{B_2} is limited by base current to zero volts, V_{C_1} is at the negative supply voltage ($-12V$) and V_{B_1} is approaching cut-off (using Fig 18b waveforms).

Instant A. V_{B_1} reaches cut-off...

... TR_1 conducts and V_{C_1} falls from $-12V$ towards its low working value...

... This fall is transferred through C_2 to TR_2 base causing V_{B_2} to fall from zero...

... The fall in V_{B_2} causes V_{C_2} to rise towards $-12V$...

... The rise in V_{C_2} is transferred through C_1 to TR_1 base causing V_{B_1} to rise further above zero and the action is *cumulative*.

An avalanche results which cuts off TR_2 and leaves TR_1 conducting. V_{C_1} falls rapidly to its working value, V_{B_1} rises above zero volts, V_{C_2} rises by the same amount and V_{B_2} falls to a positive value well beyond cut-off.

Interval A to B. C_1 charges through R_{L2} causing V_{B1} to return to zero and V_{C2} to rise towards $-12V$ with a time constant to $C_1 R_{L2}$ seconds. C_2 discharges through R_2 and V_{B2} rises towards the negative aiming voltage with a time constant of $C_2 R_2$ seconds.

Instant B. As soon as V_{B2} rises through cut-off TR_2 starts to conduct and a second avalanche takes place with the roles of TR_1 and TR_2 reversed in relation to those at instant A. This second avalanche cuts off TR_1 and leaves TR_2 conducting. V_{C2} falls rapidly to its working value, V_{B2} rises above zero volts, V_{C1} rises by the same amount and V_{B1} falls to a positive value well beyond cut-off. Thereafter the action is repeated.

The output may be taken from either collector or from both. The remarks made earlier about control of frequency and symmetry and the use of an aiming voltage are applicable to this circuit also. In addition the circuit may be synchronized by applying *negative-going* sync pulses to one or the other or both bases.

Emitter-coupled Emitter-timed Multivibrator

Steeper leading edges can be obtained, as in the cathode-coupled multivibrator, if the capacitor connected to the output electrode can be eliminated. In the cathode-coupled multivibrator V_2 anode had no coupling capacitor connected to it so that V_{a2} output had very steep leading edges. The transistor equivalent is the emitter-coupled, emitter-timed circuit shown in Fig 19a. The action may be explained with reference to Fig 19b.

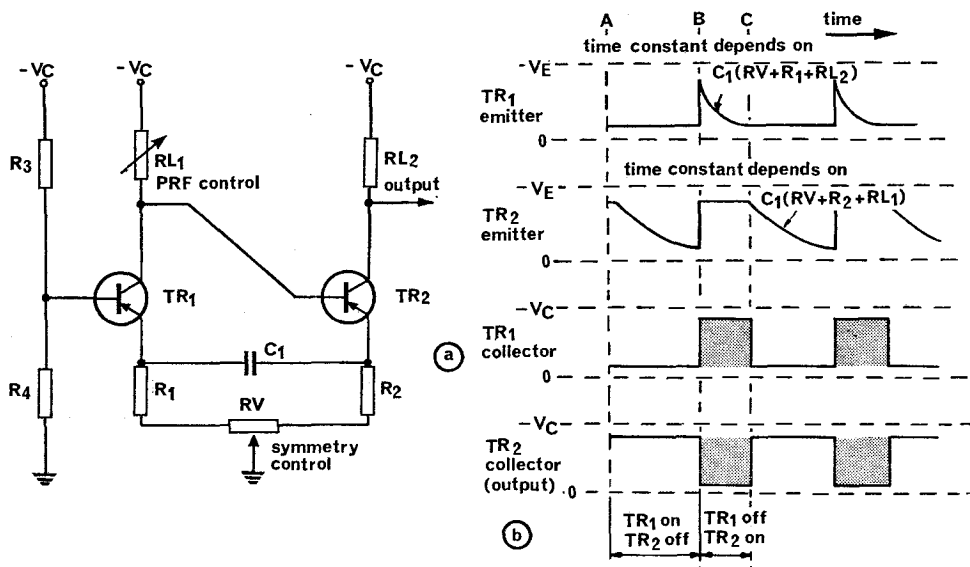


FIG 19. EMITTER-COUPLED, EMITTER-TIMED MULTIVIBRATOR, CIRCUIT AND WAVEFORMS

If we assume that TR_1 is conducting initially, its emitter and collector voltages are both at a value slightly above earth (*'negative up'* waveforms); the actual value depends upon the resistors R_1 and R_{L1} . TR_2 is then cut off because its base is at the same voltage as TR_1 collector and its emitter is held *negative* with respect to its base by the charge on C_1 . TR_2 collector is thus at the negative supply voltage level.

Interval A to B. With TR_1 conducting, C_1 commences to charge in the opposite direction through RV , R_2 and R_{L1} on a time constant of $C_1 (RV + R_2 + R_{L1})$ seconds. As it does so the voltage at TR_2 emitter falls towards earth. The collector voltages remain the same as at instant A so long as TR_1 is on and TR_2 off.

Instant B. When TR_2 emitter has fallen sufficiently, TR_2 begins to conduct, taking the charging current from R_2 which was previously passing into TR_1 emitter. This fall of current in TR_1 causes TR_1 collector voltage to *rise* towards the negative supply voltage level taking TR_2 base voltage with it. As TR_2 base becomes more negative, TR_2 conducts more heavily and a cumulative action results which causes TR_2 emitter to rise towards $-V_C$ taking TR_1 emitter voltage with it (via C_1). TR_2 collector voltage falls towards earth as TR_1 collector voltage rises towards $-V_C$.

Interval B to C. With TR_2 conducting, C_1 now commences to charge in the opposite direction through RV , R_1 and R_{L2} . As it does so TR_1 emitter voltage falls until it reaches the point where TR_1 starts to conduct. When this happens TR_1 collector voltage falls, taking TR_2 base voltage with it. TR_2 emitter is held negative by the charge on C_1 so that the fall in TR_2 base voltage cuts TR_2 off and TR_2 collector voltage rises to $-V_C$. Thereafter the cycle is repeated as from instant A.

In this circuit, variation of R_{L1} varies the time constant $C_1(RV + R_2 + R_{L1})$ and so determines the time for which TR_1 is on. It is therefore a *p.r.f.* control. In some circuits various values for C_1 may be switched in, giving various selected values of *p.r.f.*, with R_{L1} as a fine control. RV is a *symmetry* control which determines the mark-to-space ratio of the output. Very short pulse durations of the order of a few microseconds may be obtained with this circuit.

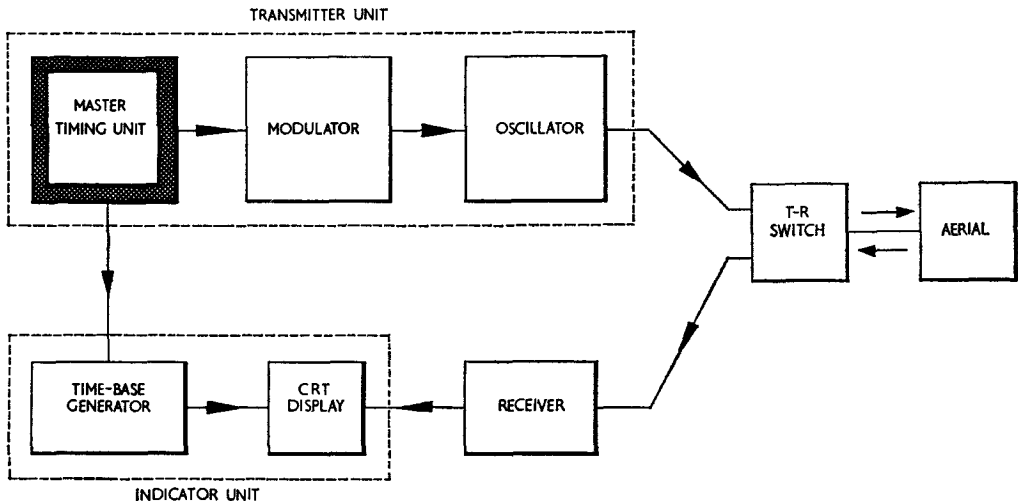


FIG 20. PRIMARY RADAR BLOCK DIAGRAM

How Multivibrators are used

Fig 20 shows the basic schematic diagram of a primary radar installation which we built up in Section 1. The master timing unit produces a waveform which triggers off each transmitter pulse and so controls the *p.r.f.* of the output. The timing waveform is also applied to the indicator c.r.t. to synchronize the spot movement with the transmitter pulses.

If we require a *p.r.f.* of 500 pulses per second we could use a multivibrator in the timing circuit with the frequency control set to give 500 cycles of output per second. To convert the square wave output from the multivibrator into timing pulses we could pass it through a differentiating CR circuit, the output from which would be a series of pips (Fig 21). The negative-going pips could then be removed by a negative limiter.

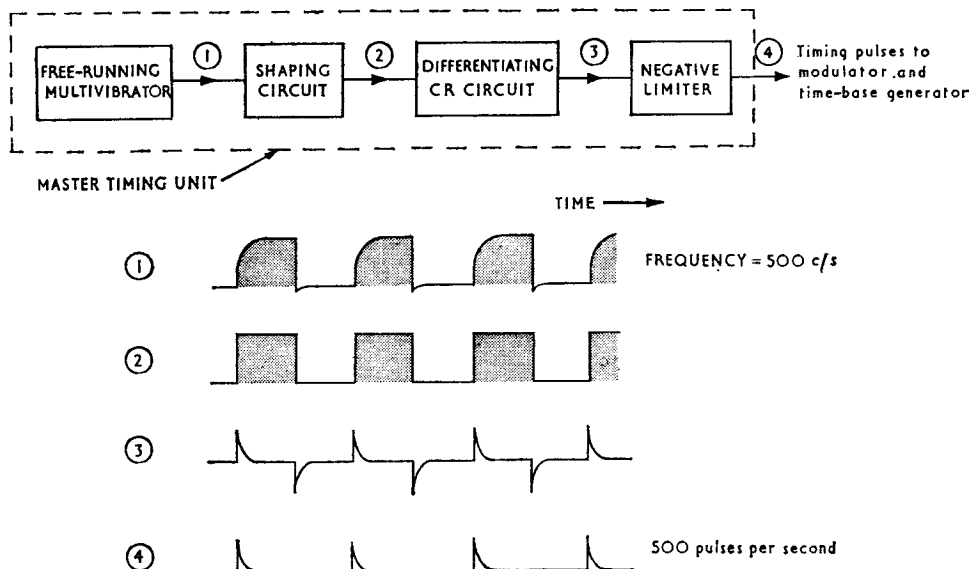


FIG 21. TYPICAL USE OF MULTIVIBRATOR

The output from the timing circuit shown in Fig 21 is a series of positive-going pips at a p.r.f. of 500 pulses per second. This waveform may be used as triggering pulses to initiate the action of other circuits or to synchronize various circuits in other parts of the installation.

CHAPTER 7

MONOSTABLE AND BISTABLE MULTIVIBRATORS

Introduction

The astable multivibrator considered in the previous chapter is a circuit which has no stable state, *ie* it is a *free-running* relaxation oscillator. Although it may be *synchronized* by applying sync pulses, it is capable of free-running with no input. The monostable and bistable multivibrators on the other hand both *require a triggering input pulse* to make them operate. We shall see how this is done in this chapter. We shall consider the monostable circuit first and then go on to the bistable circuit.

Monostable Multivibrator (Flip-flop)

Monostable multivibrators are circuits with *one stable state*. They remain in the stable state until triggered, when they then 'flip' over to the other state. They remain in the unstable state for a time decided by the circuit constants and then, *of their own accord*, 'flop' back to the original stable state. Flip-flops find many applications in radar. They may be used to generate rectangular pulses 'locked' to precise time intervals, to reshape pulse trains which have deteriorated in shape, to stretch narrow pulses into wider ones or to generate a time delay. Fig 1 illustrates how a time delay may be produced. The flip-flop produces a rectangular wave whose leading edge is coincident in time with the trigger pulse and whose trailing edge may be varied with time. The output from the flip-flop may be differentiated and then negatively limited to give a series of pulses which have a controlled variable time delay in relation to the original trigger pulses. There are a number of variations of the monostable multivibrator and in the following paragraphs we shall consider some of these basic circuits.

Anode-coupled Flip-flop

The circuit of a basic anode-coupled flip-flop and its associated waveforms are illustrated in Fig 2. The circuit is similar to that of the free-running anode-coupled multivibrator discussed in p 110. The main differences are:

a. The grid of one valve (V_1 in this circuit) is taken to a *negative bias* voltage instead of to earth or a positive aiming voltage. This bias is sufficient to hold V_1 *below cut-off* in the absence of a trigger pulse.

b. The cross-coupling impedance between V_2 anode and V_1 grid is normally a *resistance* (R_3 in this circuit). A capacitance may however be used in some circuits.

The action of the circuit may be explained as follows:

Stable state. Before the application of a trigger pulse, the circuit is in its stable state with V_1 cut off by the bias voltage applied to its grid

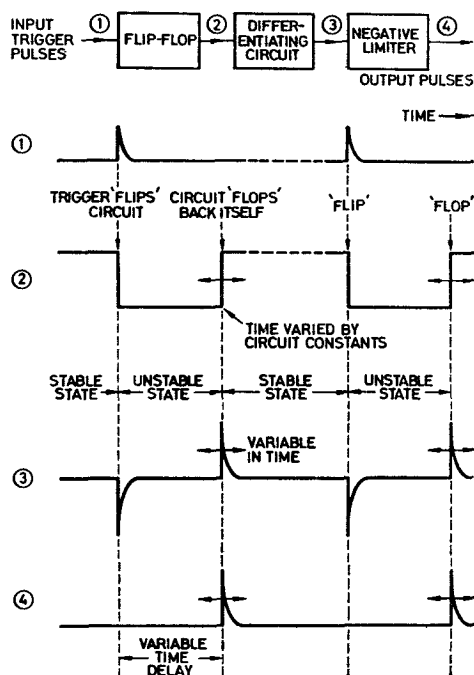


FIG 1. HOW A TIME DELAY MAY BE PRODUCED BY A FLIP-FLOP

and V_2 conducting. V_{g1} is at a negative voltage and V_{a1} at h.t. +; V_{g2} is limited to zero volts by the flow of grid current through R_2 and V_{a2} is at a low working voltage.

Instant A. The positive trigger pulse applied to V_1 lifts V_{g1} above cut-off and V_1 conducts. The usual multivibrator avalanche then takes place causing the circuit to 'flip' to its unstable state where V_1 is conducting and V_2 is cut off. V_{g1} rises to zero volts, where it is clamped by grid current limiting, and V_{a1} falls to a low working value. V_{g2} is driven below cut-off and V_{a2} rises.

Interval A to B. This is the normal multivibrator relaxation period during which V_{a2} is at a voltage just below h.t. + as determined by the potential divider R_{L2} , R_3 , R_1 . With only the valve capacitance of the stage to charge, i.e. no cross-coupling capacitor connected to V_2 anode, V_{a2} rises very quickly. C_2 discharges through R_2 towards the aiming voltage causing V_{g2} to rise exponentially.

Instant B. V_{g2} reaches cut-off and V_2 conducts. The usual avalanche then occurs and results in the circuit 'flopping' back to its original stable state, in which V_2 is conducting and V_1 cut off. V_{g2} rises to zero volts, where it is clamped by grid current limiting, and V_{a2} falls to its working value. V_{g1} is driven well below cut-off and V_{a1} rises by the same amount as V_{g2} has risen from cut-off.

Interval B to C. In this relaxation period V_{a1} rises exponentially to h.t. + as C_2 charges through R_{L1} . V_{g1} returns rapidly to its original level and remains there until the next trigger pulse arrives. The circuit can stay in this stable state indefinitely. V_{g1} is held below cut-off by the bias voltage and V_{a1} is at h.t. +; V_{g2} is limited to zero volts and V_{a2} is at its working voltage.

Instant C. On receipt of the next trigger pulse the action is repeated as from instant A.

The output may be taken either from the anode of V_2 (V_{a2}) or from the anode of V_1 (V_{a1}) depending upon whether a positive or a negative-going output is required. The output from V_2 anode has the better shape since there is no capacitor connected to this electrode. The *p.r.f.* of the circuit is determined by the repetition rate of the *trigger pulses*, one complete cycle of output being obtained for each trigger pulse. The *pulse duration* AB is determined by the time constant C_2R_2 and by the aiming voltage and may be controlled by varying R_2 . The instant of time at which the trailing edge of the pulse occurs relative to the leading edge can therefore be varied.

This circuit may also be triggered by applying *negative-going* trigger pulses to the *anode* of V_1 . Each negative pulse is applied through C_2 to V_2 grid and appears as an *amplified* and *inverted* pulse, i.e. positive-going at V_2 anode, whence it is applied to V_1 grid via R_3 to cut on V_1 .

The transistor equivalent of the circuit just discussed is known as the *collector-coupled flip-flop*. The circuit and waveforms are shown in Fig 3. The action of the circuit may be deduced from the description given for the action of the valve anode-coupled flip-flop.

Emitter-coupled Flip-flop

The basic circuit of an emitter-coupled flip-flop is illustrated in Fig 4. Here, the base of TR_2 is connected through R_3 to the negative supply $-V_C$. The base current supplied through R_3 is then such that TR_2 , in the absence of triggering, conducts heavily and the large voltage drop across the common emitter resistor R_E makes the emitters of *both* transistors *negative* with respect to earth. Since the base of TR_2 is connected through R_3 to a negative voltage *greater* than that at its emitter TR_2 conducts. The base of TR_1 however is connected to a point of fixed voltage which is only one volt negative with respect to earth; V_{B1} is therefore *positive* with respect to its emitter and TR_1 is cut off. The circuit remains in this stable state, with TR_2 on and TR_1 off, until TR_1 is made to conduct by a *negative-going* input pulse to its base.

When TR_1 conducts on application of the trigger pulse, its collector voltage V_{C1} falls from the negative supply voltage level towards zero. This change in V_{C1} is transferred instantaneously

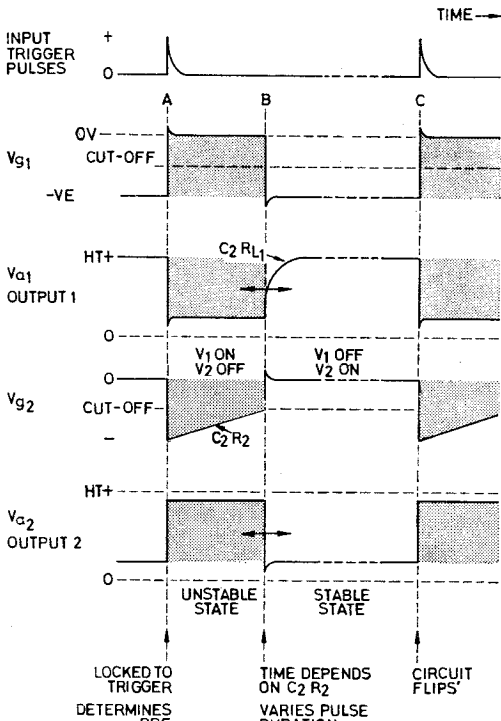
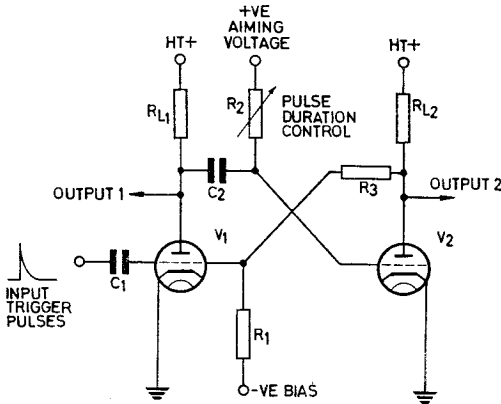


FIG 2. BASIC ANODE-COUPLED MONOSTABLE MULTIVIBRATOR AND ASSOCIATED WAVEFORMS

to TR_2 base through C_2 causing TR_2 to conduct less heavily. V_{C2} therefore rises towards $-V_C$, and the common emitter voltage V_E falls towards earth. (This occurs because the fall in current through R_E due to TR_2 is greater than the rise due to TR_1 since the change of voltage at TR_2 base has been amplified through TR_1). The fall in V_E causes TR_1 emitter voltage to become more positive with respect to its base and TR_1 conducts more heavily. The usual avalanche occurs to 'flip' the circuit into its unstable state in which TR_2 is off and TR_1 on.

TR_2 is held cut off by the charge on C_2 applied to its base and the unstable state is maintained during the time that C_2 is discharging through R_3 to the aiming voltage $-V_C$. This causes V_{B2}

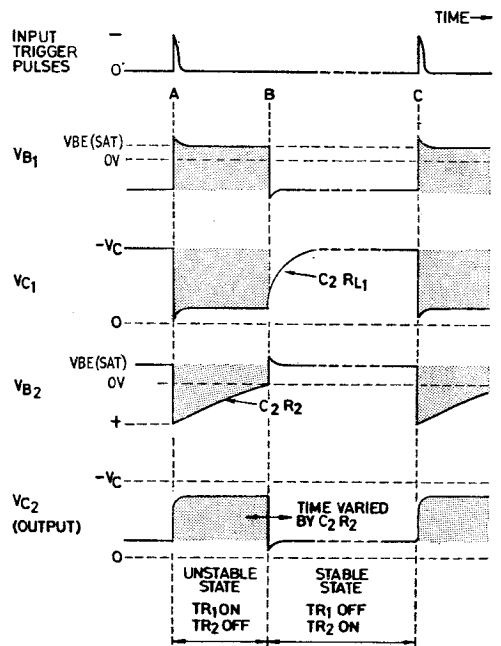
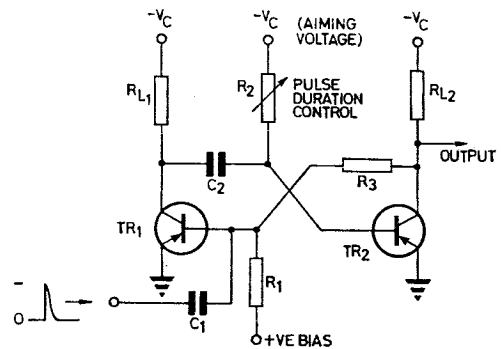


FIG 3. COLLECTOR-COUPLED MONOSTABLE MULTIVIBRATOR, CIRCUIT AND WAVEFORMS

to rise exponentially towards cut-on. After a time determined mainly by the time constant C_2R_3 , V_{B2} reaches cut-on and TR_2 conducts. A second avalanche is thus initiated and the circuit 'flops' back to its original stable state in which TR_2 is on and TR_1 off. It will remain in this state until the next trigger pulse is applied to TR_1 base.

We can see from Fig 4 that a very good negative-going rectangular wave is obtained from TR_2 collector (V_{C2}). This is used as the output waveform. The p.r.f. is determined by the trigger rate, and the pulse duration by the time constant C_2R_3 (and the aiming voltage). Variation of R_3 varies the pulse duration (or delay).

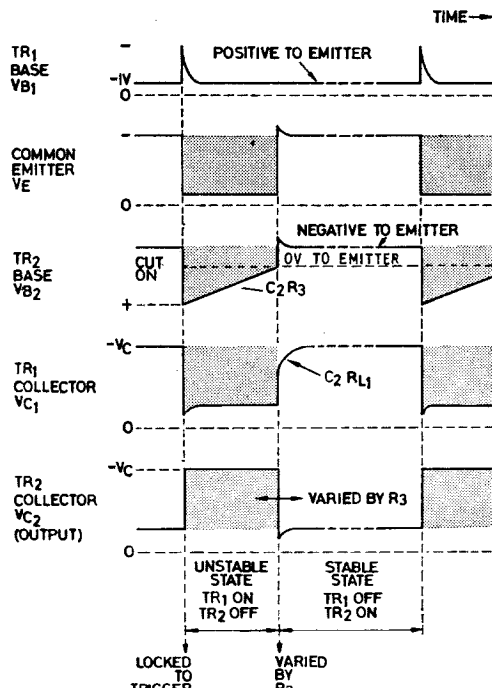
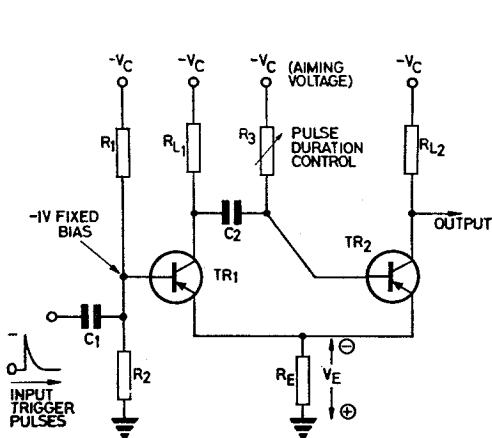


FIG 4. EMITTER-COUPLED MONOSTABLE MULTIVIBRATOR, CIRCUIT AND WAVEFORMS

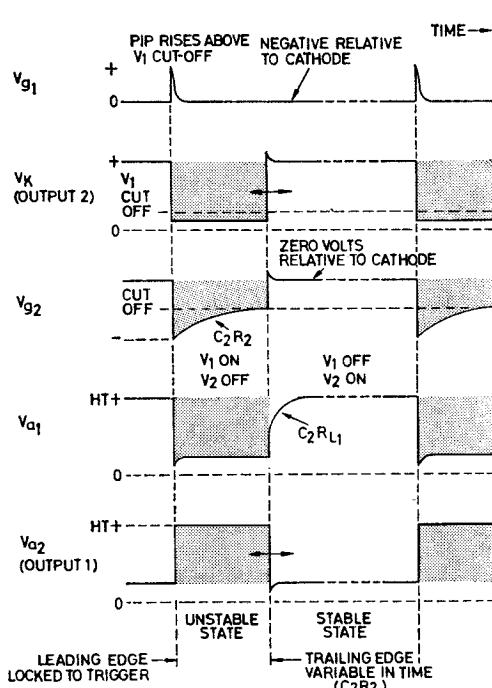
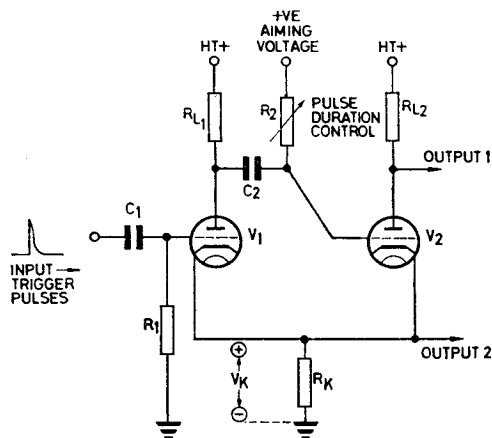


FIG 5. CATHODE-COUPLED FLIP-FLOP, CIRCUIT AND WAVEFORMS

The valve equivalent of this circuit is the *cathode-coupled flip-flop*, the circuit and waveforms of which are illustrated in Fig 5. Examination of Figs 4 and 5 will show that the two circuits are similar, so that the action of the valve version may be deduced from the previous paragraphs. An output is sometimes taken from the cathode (output 2); this is a low impedance source and is useful for matching to other stages. Like the anode-coupled flip-flop, this circuit may be triggered by applying *negative-going* trigger pulses to V_1 anode.

Bistable Multivibrator

The bistable trigger or toggle is a two-state circuit in which *both states* are stable. In each stable state one of the two transistors (or valves) is cut off and the other is conducting, and the circuit is capable of remaining indefinitely in either stable state. To change over from one state to the other the circuit must be suitably triggered and, to go through a full cycle, two triggering pulses—one to each stage—are needed.

The bistable multivibrator has many applications: it can be used as a frequency divider, as part of a counting circuit or as a storage device in a computer (see p 584 of AP 3302, Part 1B). We shall see something of these applications later in these notes. The bistable multivibrator has been given many names, but the name most commonly used is the "*Eccles-Jordan*" circuit. Let us now examine the basic Eccles-Jordan circuit using valves.

Eccles-Jordan Bistable Circuit

The basic circuit of an Eccles-Jordan bistable multivibrator is shown in Fig 6a. This circuit differs from the anode-coupled astable multivibrator in two respects:

- The cross-coupling impedances are *resistors* so that they provide d.c. coupling between anode and grid of opposite valves.
- The grid resistors of *both* valves are returned to a negative bias point.

R_{L1} , R_1 , R_2 and R_{L2} , R_3 , R_4 each act as a potential divider chain connected between h.t. + and the negative bias. Thus each grid voltage lies somewhere between the negative bias and h.t. + voltage levels. It is usually arranged, by correct selection of bias and resistor values, that the maximum grid voltage of the conducting valve is zero when the other valve is cut off. Thus in Fig 6b, if V_1 is cut off, the three resistors each have 100V dropped across them and the division of voltage is as shown. The voltage division in the potential divider connected to the anode of the conducting valve is, of course, different (Fig 6c). If V_2 is conducting, its anode current will produce an *additional* voltage drop

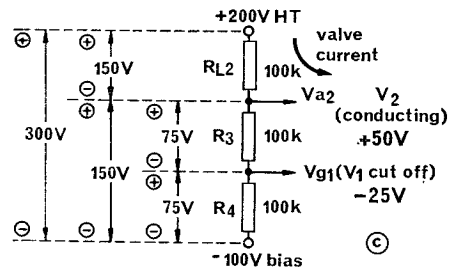
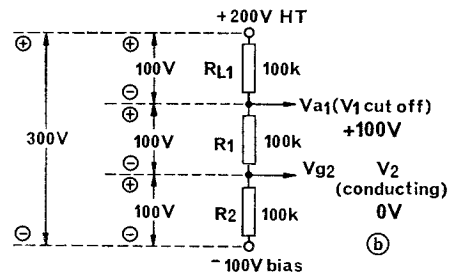
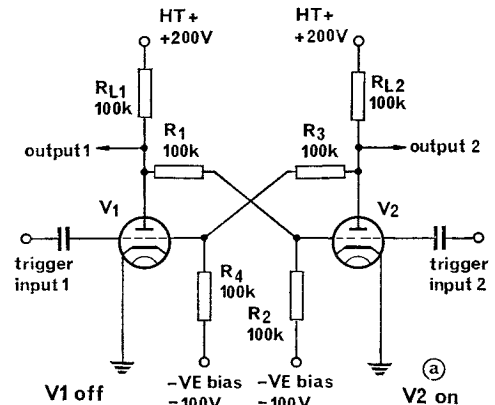


FIG 6. ECCLES-JORDAN BISTABLE MULTIVIBRATOR, CIRCUIT ARRANGEMENTS

across R_{L2} . If this is sufficient to bring V_{a2} down to $+50V$ then across R_{L2} we have a voltage drop of $150V$. We thus have the remainder, *ie* $150V$, divided equally between R_3 and R_4 ($75V$ across each) and V_{g1} is held at $-25V$ with respect to earth, sufficient to hold V_1 cut off.

This represents one stable state in which V_1 is cut off and V_2 is conducting. This state can be held indefinitely because there are no capacitors in the circuit to discharge and lift V_{g1} to the cut-off point, *ie* there is no true 'relaxation period' in this circuit. To switch over to the other stable state we can apply either a *positive* pulse to V_1 grid to *cut on* V_1 or a *negative* pulse to V_2 grid to *cut off* V_2 . Of the two methods the application of a negative pulse to the conducting valve is more common because a larger triggering voltage is usually required to cut on a non-conducting valve.

If we apply a negative trigger pulse to the grid of the conducting valve V_2 , its anode current falls and V_{a2} rises. This causes V_{g1} to rise and V_1 conducts. An avalanche thus occurs to switch the circuit over to its other stable state in which V_1 is conducting and V_2 is cut off. Again it will remain in this state *until triggered out of it*. This is done by applying another negative trigger pulse, to the grid of V_1 this time. The waveforms for this circuit are shown in Fig 7. Note that the p.r.f. and the pulse duration are *both* determined by the trigger pulses.

An Eccles-Jordan circuit which avoids the need for a separate bias supply is shown in Fig 8. In this circuit, either V_1 or V_2 is conducting during the stable states and the resulting current through R_K provides the appropriate bias voltage. The capacitor C_K ensures that the d.c. bias is not affected when switching from one stable state to the other.

The circuit and waveforms of an Eccles-Jordan circuit using transistors are shown in Fig 9.

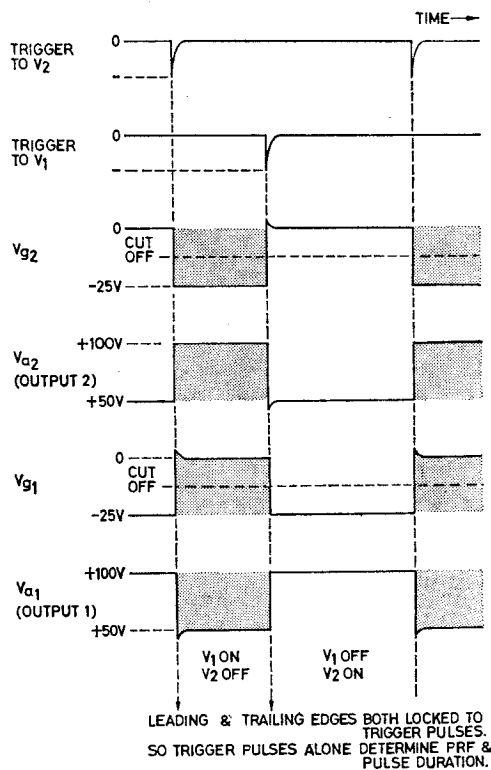


FIG 7. TYPICAL WAVEFORMS, ECCLES-JORDAN BISTABLE CIRCUIT

The action of this circuit is virtually the same as that just described for the valve version and may be deduced from the remarks of the previous paragraphs.

Note that the pips on the output waveforms are the result of grid current or base current limiting at the input. If the valve or transistor is worked under *bottoming* conditions the pips at the grid or the base will not be *reflected* in the output waveforms.

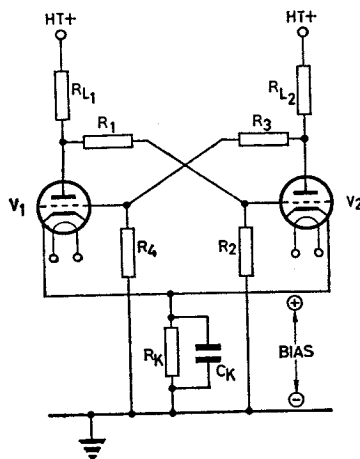


FIG 8. CATHODE BIAS IN AN ECCLES-JORDAN CIRCUIT

So far we have assumed that the output electrode has no capacitance connected to it so that the voltage at the output electrode can rise and fall very rapidly, giving steep-sided outputs. In practice, the grid-cathode input capacitance (in the valve version) and the base-emitter input capacitance (in the transistor version) will charge and discharge via the appropriate cross-coupling resistance in accordance with the voltages acting on the capacitances. This charge and discharge obviously takes *time* and the result is that the switching speed is limited and the output waveforms are not so steep-sided. To reduce these effects, Eccles-Jordan circuits often have small capacitors *in parallel* with the cross-coupling resistors (Fig 10). These 'speed-up' or 'commutating' capacitors cause faster switching because any voltage change at the anode (or collector) is *immediately* transferred to the input of the opposite stage.

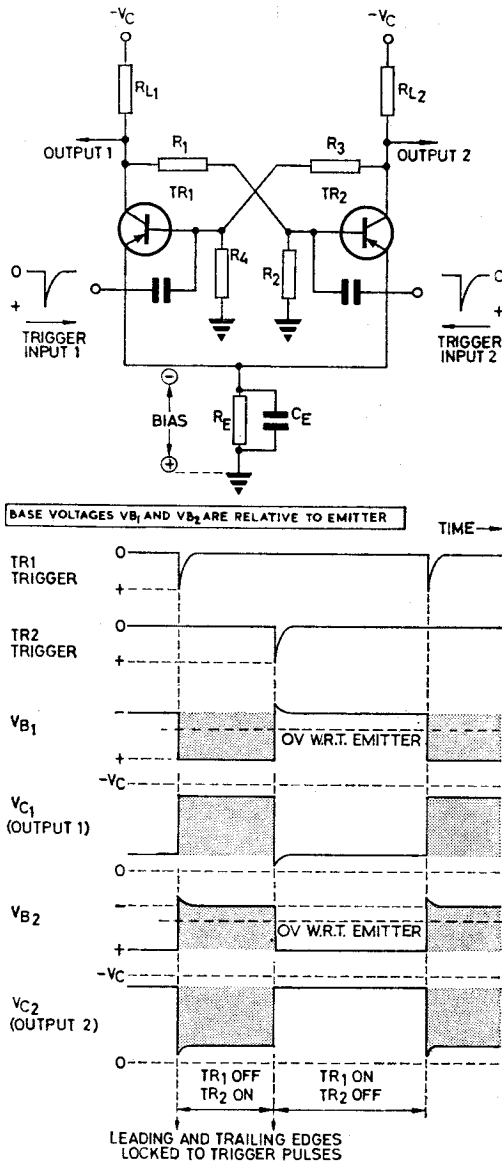


FIG 9. TRANSISTOR BISTABLE CIRCUIT AND ASSOCIATED WAVEFORMS

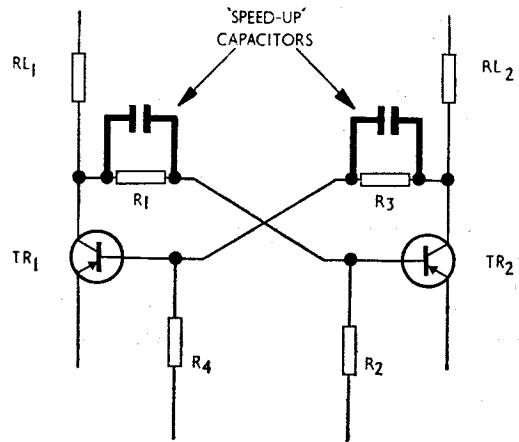


FIG 10. INSERTION OF COMMUTATING CAPACITORS TO IMPROVE SWITCHING SPEED

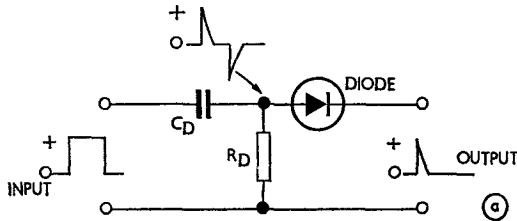


FIG 11.

TYPICAL TRIGGER CIRCUITS

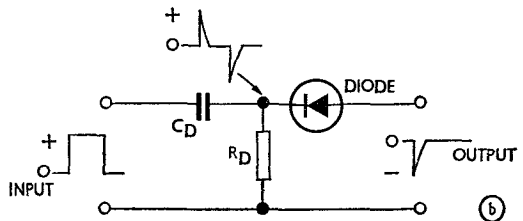


FIG 12.

Triggering Arrangements

In many switching applications the input pulse to a stage is a rectangular wave as shown in Fig 11. If this input is applied to a short CR differentiating circuit the familiar positive and negative pips are produced. If the differentiated waveform is then fed through a diode connected as shown in Fig 11, only the positive-going spikes appear at the output. This acts as the trigger pulse. By reversing the diode, as in Fig 12, a negative-going trigger pulse is obtained.

If we now connect the trigger circuit of Fig 11 to an Eccles-Jordan bistable circuit, as shown in Fig 13, we can cause it to switch between stable states. If TR_1 is on and a switching pulse is applied to the terminal marked 'S', the resulting positive spike at the base of TR_1 initiates the change-over to the state where TR_1 is off and TR_2 on. Once TR_1 is off, further trigger pulses to 'S' have no effect; they merely tend to drive TR_1 further off. To revert to the initial stable state we must apply a pulse to the terminal marked 'R'. The resulting positive spike at the base of TR_2 initiates the change-over to the first stable state where TR_1 is on and TR_2 off. The labels 'S' and 'R' stand for 'set' and 'reset' respectively. We can therefore set the bistable by a pulse to 'S' and reset it by a similar pulse to 'R'. Either of the two states may be selected at will. This has many applications in computers and counting circuits.

In some circuits we have to apply the trigger pulses from a *single source* to both stages simultaneously. This may be done with the arrangement shown in Fig 14. A series of pulses, all of the same kind, will then cause the circuit to switch between one state and the other. When a pulse arrives it finds one transistor off and the other on. The positive trigger pulses have no effect on the 'off' transistor but they do effect the other transistor, initiating a change-over. The next pulse finds the positions reversed and switches the circuit back to its first state.

The same reasoning applies to valve circuits except that the trigger pulses there are usually *negative-going*.

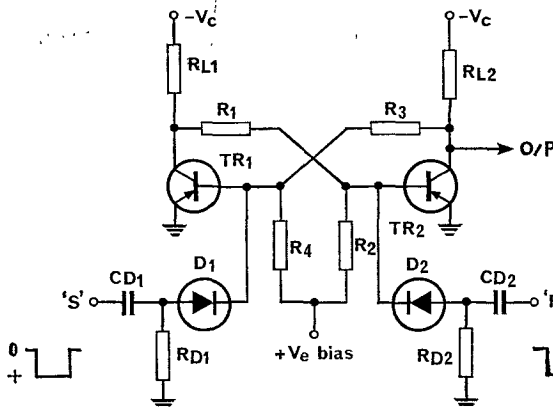


FIG 13. 'SET' AND 'RESET' OF AN ECCLES-JORDAN BISTABLE CIRCUIT

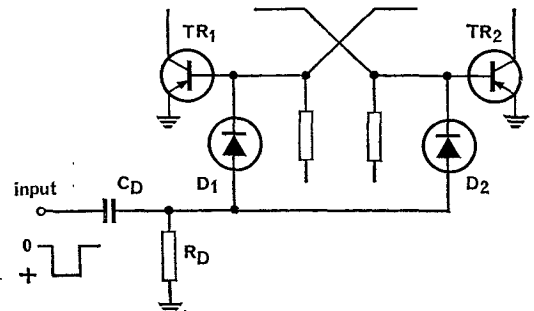


FIG 14. CONTINUOUS SWITCHING OF AN ECCLES-JORDAN CIRCUIT BY A PULSE TRAIN

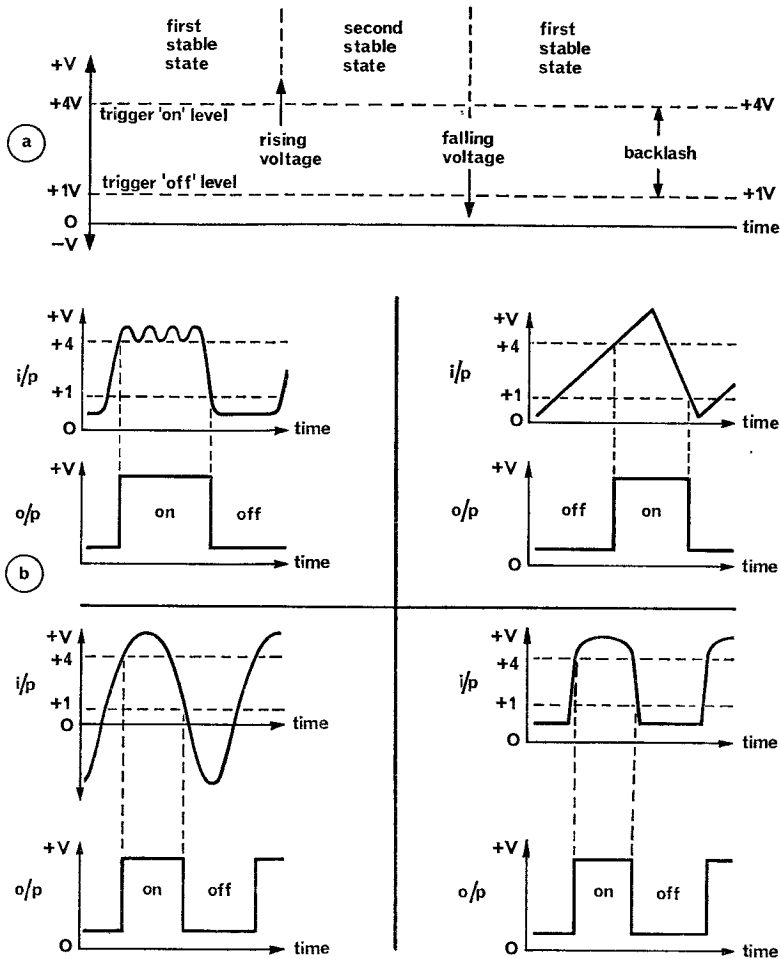


FIG 15. SWITCHING IN A SCHMITT TRIGGER CIRCUIT

Schmitt Trigger Circuit

The Schmitt trigger circuit is, in effect, a cathode-coupled version of the Eccles-Jordan bistable multivibrator. The triggering arrangements, however, are different. In the Schmitt trigger the output voltage takes up one or the other of its two stable levels depending upon whether the input voltage is *above* or *below* some chosen value. The circuit is usually designed so that the change-over between stable states occurs for *different* values of input. For example, to change from the first stable state to the second, the input may have to *rise through* +4V. To switch back to the first stable state the input may have to *fall below* +1V (Fig 15a). The *difference* between the two triggering levels (3V in this example) is known as the '*backlash*' or '*hysteresis*' of the system.

For a bistable circuit which operates with this type of triggering input, a good square wave output is obtained for any input which passes through the required triggering points. This is illustrated in Fig 15b. In addition, the Schmitt trigger may be used to determine when the amplitude of a signal applied to it has exceeded a critical value (the 'trigger-on' level).

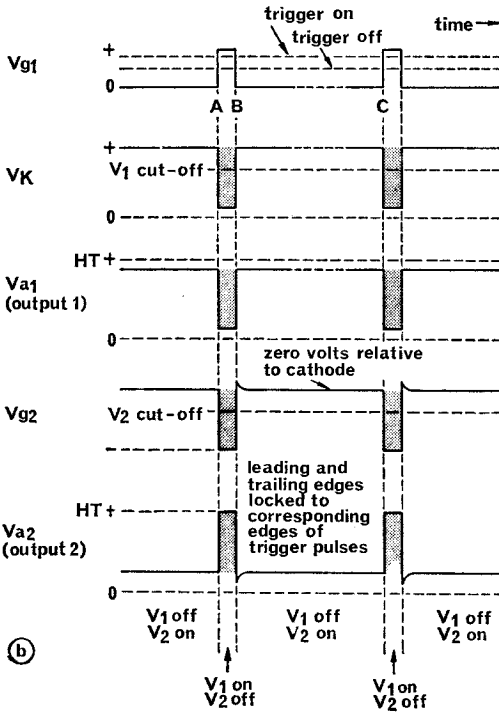
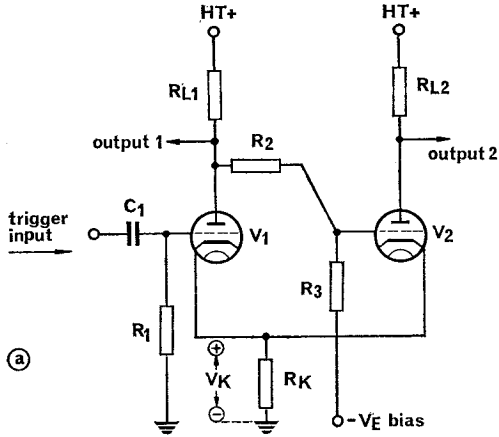


FIG 16. SCHMITT TRIGGER AND ASSOCIATED WAVEFORMS

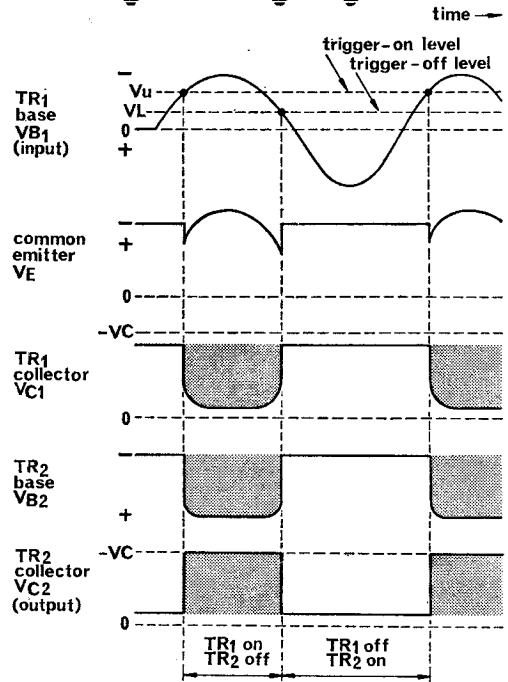
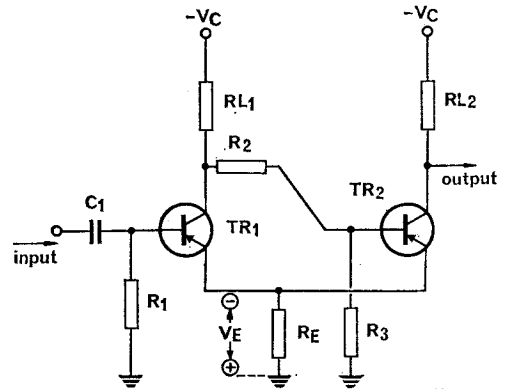


FIG 17. TRANSISTOR SCHMITT TRIGGER, TYPICAL APPLICATION

The basic circuit of a Schmitt trigger using valves is shown in Fig 16a, with associated waveforms in Fig 16b. Under normal conditions, with no trigger input applied to V_1 , V_2 is conducting heavily and the resulting voltage drop across R_K raises the cathode voltage V_K of both valves above earth. With V_{s1} at zero volts, V_K is sufficiently positive with respect to V_{s1} to keep V_1 cut off. V_{a1} is thus at a high voltage (determined by the values of R_{L1} , R_2 , R_3 and the h.t. and bias voltages). V_{a2} is also determined by the potential divider connected between h.t.+ and the negative bias point. These are adjusted such that V_{a2} tends to go positive but is held at zero volts relative to its cathode by grid current limiting. V_{a2} is therefore at its working voltage. This is the first stable state, in which V_1 is held cut off and V_2 is conducting.

Instant A. When a positive trigger is applied to V_1 a point is reached (the 'trigger-on' level) where the cathode bias due to R_K is overcome and V_1 conducts. V_{a1} therefore falls, V_{g2} falls and, because the fall in V_{g2} has been amplified by V_1 , the decrease in current through R_K due to V_2 is greater than the increase due to V_1 , so that V_K also *falls*. The fall in V_K increases V_1 conduction, initiating an avalanche which rapidly switches the circuit over to its other stable state in which V_1 is on and V_2 off.

Interval A to B. The circuit remains in this stable state in which V_{a2} is at h.t. + and V_{a1} is at a low working value.

Instant B. The trailing edge of the trigger pulse reduces V_{g1} , causing V_{a1} to rise. V_{g2} therefore rises and when the trailing edge of the trigger pulse has fallen to the 'trigger-off' level, V_{g2} rises above cut-off and V_2 conducts. Since the rise in V_{g2} has been amplified by V_1 , V_2 anode current increases by more than V_1 anode current decreases, so that V_K also *rises*. The rise in V_K reduces V_1 conduction still further and a second avalanche occurs which rapidly switches the circuit back to its original stable state in which V_1 is off and V_2 on.

Interval B to C. The circuit remains in this stable state until another trigger pulse applied to V_1 raises V_{g1} sufficiently to overcome the cathode bias V_K (at instant C).

With the arrangement shown in Fig 16 the circuit is being triggered rapidly from one stable state to the other by small positive-going trigger voltages applied to V_1 grid. (Larger *negative-going* triggers to V_2 grid produce similar results). At each avalanche an amplified trigger voltage appears at both anodes (in anti-phase with each other) and by making R_K variable the input voltage necessary to start an avalanche can be made very small, e.g. 0.1V. Like the Eccles-Jordan circuit, a speed-up capacitor may be shunted across R_2 . Since, in this application, the leading and trailing edges of the output pulse are locked to the corresponding edges of the trigger pulse, we are effectively converting a weak input pulse into a larger amplitude output pulse of the same p.r.f. and pulse duration.

A transistor version of the Schmitt trigger is illustrated in Fig 17. The action is somewhat similar to that just described for the valve version, but for p-n-p transistors, polarities are reversed.

The resistor values are so chosen that when no voltage is applied to the input, TR_1 is off and TR_2 on. If a voltage more negative than V_U (the 'trigger-on' level) is applied to the input, TR_1 cuts on and TR_2 cuts off. So long as the input is then held more negative than V_L (the 'trigger-off' level) TR_1 remains on and TR_2 off.

When the input falls below V_L towards zero, TR_1 cuts off again and TR_2 conducts. Thus the state in which the circuit is operating depends upon whether the input is above V_U or below V_L .

In the application shown in Fig 17, the Schmitt trigger is being used as a *squarer*. By adjusting the mean input level (zero in Fig 17) we can produce a symmetrical square wave output from a sine wave input. The Schmitt trigger also has many other useful applications.

OTHER SQUARE WAVE GENERATORS

Introduction

Although the majority of pulse generators are based on one of the family of multivibrators discussed in the preceding chapters there are occasions when other circuits capable of producing a rectangular wave output are preferred. Many of these circuits use only one stage and are used mainly for special applications. We shall consider some of these circuits in this chapter.

Complementary Regenerative Pair Circuit

This circuit uses transistors of *opposite* polarity—one p-n-p and other n-p-n—and has no valve equivalent. One version of the circuit is illustrated in Fig 1. It is, in effect, a monostable or flip-flop circuit. In this arrangement TR_1 is a p-n-p transistor and TR_2 a n-p-n type. TR_1 therefore requires a *negative* voltage to its base to cut it on, and TR_2 base must be *positive* with respect to its emitter before it conducts. In the quiescent condition *both* transistors are off. Before the application of a trigger pulse, with no current to its base, TR_1 is cut off and its collector voltage is at $-V_C$. There is therefore no drive voltage to the base of TR_2 which is also cut off. Under these conditions C_2 is discharged and the output is at earth. This is the stable state.

If we apply a *negative* trigger pulse to TR_1 base, TR_1 cuts on and its collector voltage falls towards earth because of the voltage drop across R_{L1} . (Note that we are using the '*negative up*' convention in this circuit.) TR_2 base is therefore driven down and becomes *positive* with respect to its emitter so that TR_2 also cuts on. TR_2 collector voltage then rises towards $-V_C$ because of the voltage drop across R_{L2} , and the output also rises. This rise is applied through C_2 (which cannot charge instantaneously) to the base of TR_1 where it drives TR_1 hard on. An avalanche therefore occurs which '*flips*' the circuit into the state where *both* transistors are conducting. This is the unstable state in which TR_1 collector voltage is steady at just above zero volts and TR_2 collector voltage is steady at just below $-V_C$. The duration of the unstable state depends upon the time taken for C_2 to charge.

With TR_1 and TR_2 both conducting, C_2 is connected across the supply and commences

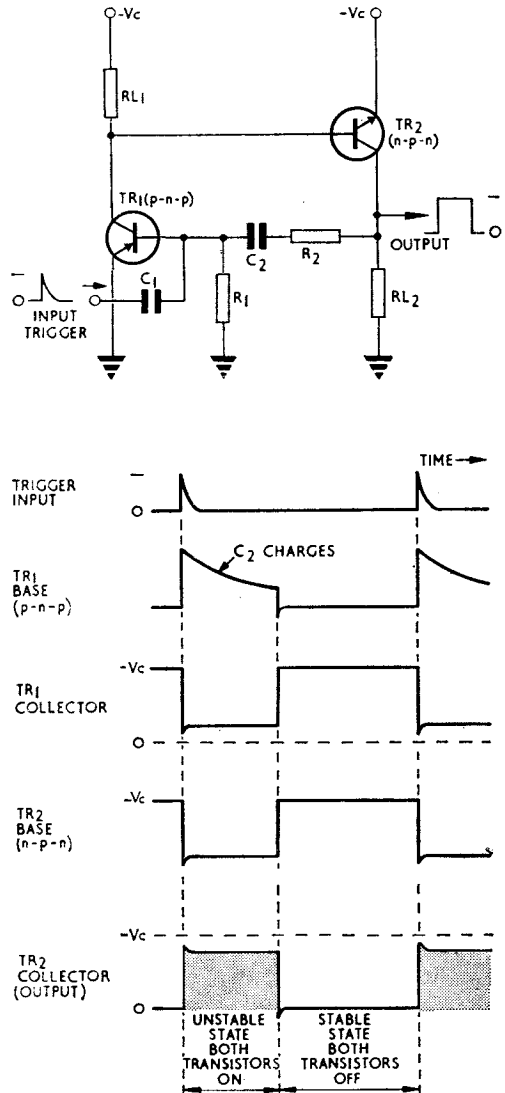


FIG 1. COMPLEMENTARY REGENERATIVE PAIR, CIRCUIT AND WAVEFORMS

to charge through the resistors connected to it. After a certain time the charge on C_2 is such that TR_1 base falls to the level where TR_1 cuts off. The circuit then 'flops' back to its original stable state in which both transistors are off.

This circuit has the advantage that in the stable state both transistors are off and the circuit therefore requires *no power* until it is triggered. It can also handle large load currents and can sustain long duration pulses.

Transitron

The transitron, the circuit and waveforms of which are illustrated in Fig 2, is sometimes referred to as a *one-valve flip-flop*. It is a monostable circuit producing a square wave output whose p.r.f. is determined by that of the trigger pulses and whose duration may be varied.

The circuit depends for its action on the fact that the total *space* current in a pentode is determined by the control grid voltage V_{g1} , but the *division* of this current between anode and screen is determined by the *suppressor* grid voltage V_{g3} . Thus if V_{g1} remains constant:

- When V_{g3} rises the anode current I_a rises and the screen current I_s falls.
- When V_{g3} falls, I_a falls and I_s rises.

Neglecting any control grid current which may flow, the sum of I_a and I_s is at all times equal to the total space current.

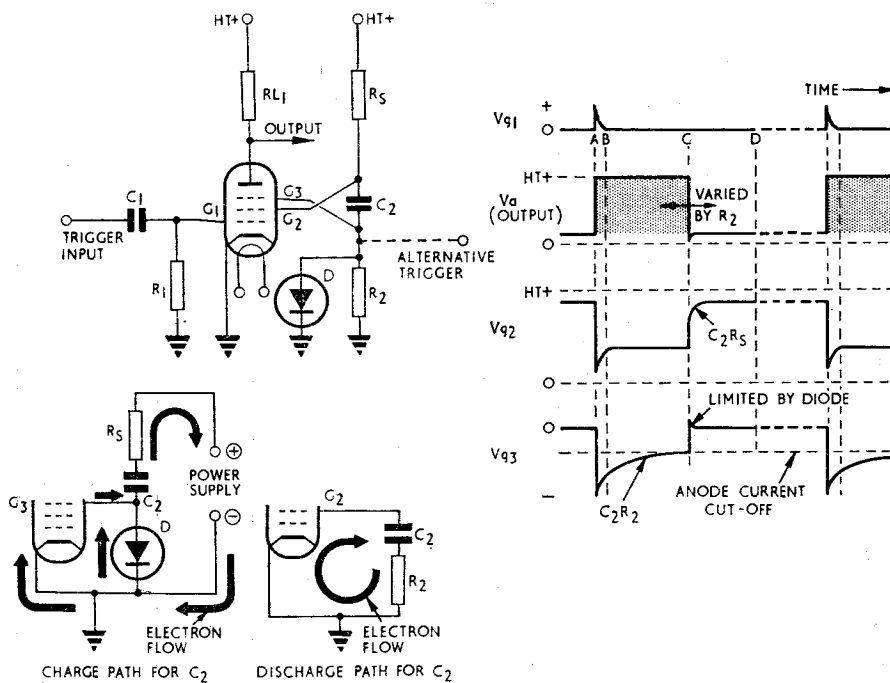


FIG 2. BASIC TRANSITRON, CIRCUIT AND WAVEFORMS

Before the application of a trigger pulse, V_{g1} is at zero volts and the valve is conducting normally with its space current shared between anode and screen. The anode voltage V_a is therefore at its low working value and the screen voltage V_{g2} is below h.t.+ by an amount determined by R_s and the screen current I_s . C_2 is charged to the screen voltage so that the suppressor grid voltage V_{g3} is zero.

Instant A. The positive trigger causes V_{g_1} to rise. The total space current therefore rises and the increased screen current I_s causes V_{g_2} to *fall* due to the increased voltage drop across R_s . Since C_2 cannot discharge instantly this fall in V_{g_2} drives V_{g_3} negative via C_2 and I_a falls, causing V_a to *rise*. Since I_a has fallen, I_s must have *risen* by the same amount (since $I_a + I_s = \text{space current}$) and this causes V_{g_2} to *fall further*. There are here all the conditions for an avalanche so that as soon as the positive trigger is applied, I_a cuts off and all the space current goes to the screen. V_{g_2} therefore falls to a low value and V_a rises to h.t. +.

Instant B. When the trigger input ceases and V_{g_1} returns to zero the total space current falls. This causes the screen current I_s to fall so that V_{g_2} and V_{g_3} both rise. However, so long as the discharge current of C_2 keeps V_{g_3} beyond the anode current cut-off point, no other change occurs.

Interval B to C. The circuit is now in its unstable state and remains there whilst V_{g_3} returns towards zero with the discharge of C_2 through R_2 (see Fig 2).

Instant C. V_{g_3} rises above the anode current cut-off point and I_a flows, causing V_a to fall. With the rise in I_a , I_s falls and V_{g_2} rises. The rise in V_{g_2} is transferred through C_2 to the suppressor grid and V_{g_3} rises further. I_a therefore rises further, I_s falls still more and an avalanche results which continues until V_{g_3} is sufficiently positive to cause the flow of *suppressor grid current*. V_a falls to its low working value and V_{g_2} rises.

Interval C to D. C_2 now recharges rapidly through R_s (see Fig 2). As it does so, V_{g_3} returns to zero and V_{g_2} returns to its original level. The circuit is now in its original stable state and *remains there* until the next trigger pulse is applied.

The square wave output is taken from the anode. It is locked to the trigger pulses in time and its pulse duration is determined by the time taken for the suppressor grid voltage to rise to the anode current cut-off level, i.e. by the time constant $C_2 R_2$. By varying R_2 , the pulse duration is varied. The diode D is inserted to limit the positive rise of V_{g_3} at instant C.

This circuit may also be triggered by applying a *negative* pulse to the *suppressor grid*. At instant A the negative pulse reduces I_a , causing I_s to rise. V_{g_2} and hence V_{g_3} fall so that I_a falls further. An avalanche results and the circuit action is then as previously described.

Phantastron

This circuit is developed from the Miller timebase circuit which we shall consider in detail in a later chapter. The sawtooth waveform produced at the anode of a phantastron may be used as a timebase waveform. More usually, however, it is the square wave output from either the screen grid or the cathode which is required. The basic circuit of a phantastron and its associated waveforms are shown in Fig 3. In this circuit R_1 is the grid resistor to limit the grid to zero volts when grid current flows, and R_L is large enough to cause *bottoming* at a certain low value of V_a . The most distinguishing feature of this circuit is the 'Miller' capacitor C_1 connected between anode and control grid. Note also that neither R_s nor R_K is decoupled; rapid voltage changes can therefore occur at the screen grid and the cathode.

In the stable state, before the application of a trigger pulse to the suppressor grid, the valve is conducting normally. The grid voltage V_{g_1} is held at zero volts relative to its cathode by grid current limiting via R_1 . The cathode voltage V_K (and hence V_{g_1}) is therefore some volts *positive to earth* because of the voltage drop produced across R_K by the space current. The suppressor grid is connected to *earth* through R_2 and so is *negative* with respect to its cathode. The value of R_K is made large enough to ensure that the suppressor grid is sufficiently negative to its cathode to *cut off the anode current*. The total space current of the valve therefore flows to the screen grid and the screen grid voltage V_{g_2} is low. With no anode current, V_a is at h.t. + and the capacitor C_1 is charged.

Instant A. The *positive* trigger pulse applied to the suppressor grid reduces the bias between the suppressor grid and the cathode sufficiently to allow anode current to flow. This causes V_a to fall and, since a capacitor cannot change its charge instantaneously, the fall of V_a is transferred to the control grid through C_1 . V_{g1} therefore falls and the resulting fall in the total space

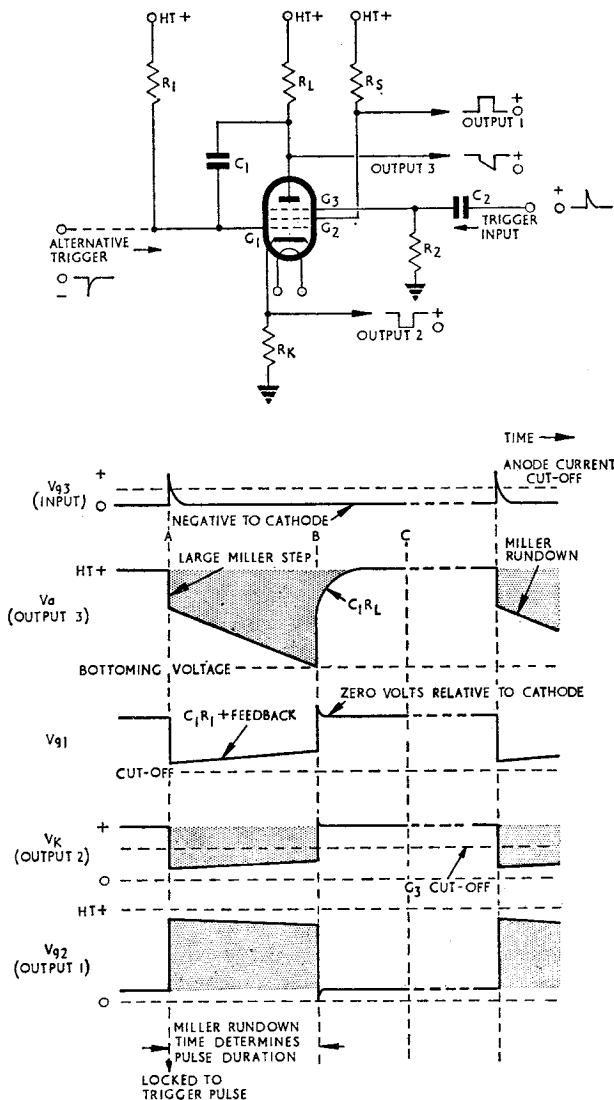


FIG 3. PHANTASTRON, CIRCUIT AND WAVEFORMS

current causes V_K to fall. The fall in V_K reduces the bias between the suppressor grid and the cathode even more and the anode current rises further. V_a therefore falls further and the action is cumulative and continues until V_{g1} is *just above* cut-off. It cannot continue beyond this point because if V_{g1} went below cut-off, V_a would rise again, lifting V_{g1} above cut-off, causing V_a to fall, and so on. The cumulative action therefore ceases with V_{g1} *just above* cut-off. Since the valve has now almost cut itself off the total space current is very small and V_K is just above zero

volts. With very little space current (and most of what there is going to the anode) the screen current is very small so that V_{g2} rises instantly to just below h.t.+. As V_K falls towards earth, the bias on the suppressor grid is reduced and more of the available space current reaches the anode. As I_a increases in this way, V_a falls. It can fall only by a limited amount however, as explained above—in this circuit by an amount equal to the grid base of the valve *plus* the fall of V_K . This sudden fall of V_a is the so-called ‘*Miller step*’ which we shall discuss in more detail later in these notes. In this circuit the Miller step is larger than normal because of the cathode follower action of V_K .

Interval A to B. V_{g1} now commences to rise as C_1 discharges through R_1 . As V_{g1} rises, the space current rises and so does V_K . The increased space current *tends* to cause V_a to fall and this fall of V_a is transferred through C_1 to the grid to *counteract* the rise of V_{g1} . This effect is known as ‘*Miller feedback*’. The feedback cannot *stop* the rise of V_{g1} completely however because, owing to the fact that V_K is also rising, the negative bias on the suppressor grid is increasing and less of the available space current is reaching the anode. The effect of the Miller feedback is to ‘*slow down*’ the rise of V_{g1} , and V_{g1} rises *linearly*. Since V_{g1} is rising linearly, V_a falls in a slow and linear manner. This variation of V_a is termed the ‘*Miller rundown*’. The rise of V_{g1} increases the screen current and V_{g2} falls slightly during the Miller rundown.

Instant B. V_a reaches its *bottoming* voltage and with V_a held at this value there is no further Miller feedback through C_1 to the grid so that V_{g1} now rises quickly. Owing to the cathode follower action, V_K also rises quickly, applying sufficient negative bias to the suppressor grid to reduce the anode current. V_a therefore rises, this rise being fed back via C_1 to the grid to cause V_{g1} to rise further. The resulting increased space current increases V_K , I_a falls still more and the cumulative action quickly cuts off I_a . The total space current is thus diverted to the screen so that V_{g2} falls to a low value as V_a rises.

Interval B to C. The capacitor C_1 now recharges quickly to its original value via R_L and the grid-cathode path of the valve, and V_a rises exponentially to h.t.+ as C_1 charges. The circuit is now in its initial stable state and will *remain in this condition* until the arrival of the next trigger pulse.

This circuit may also be triggered by applying a *negative* trigger pulse to the *control grid* and connecting the suppressor grid directly to earth. This reduces V_{g1} so that the space current falls and so does V_K . This makes the suppressor grid less negative with respect to its cathode and anode current starts to flow. The action is then the same as from instant A.

Since the avalanche which ends the square wave and returns the circuit to its stable state occurs when the anode voltage bottoms (instant B), the pulse duration of the square wave output from the screen grid depends upon the *Miller rundown time*. This can be varied by modifying the basic circuit to include ‘*anode-catching diodes*’ as shown in Fig 4:

- a. The starting point of the rundown may be adjusted as in Fig 4a. The lower the setting of V_{R1} the lower is the initial anode voltage and the less is the rundown before the bottoming voltage is reached. V_a cannot return to a higher value than that set by V_{R1} because as soon as it tends to do so the diode conducts to hold V_a at the level set by V_{R1} .
- b. An artificial bottoming voltage may be provided by stopping the anode voltage fall at any desired point by using a limiting diode as shown in Fig 4b. With the diode connected as shown, V_a can only fall until the voltage across the diode is such that it will conduct. This level is set by V_{R2} .

Sanatron

There are several forms of this circuit. A typical example with waveforms is shown in Fig 5. It produces an output similar to that of a phantatron but is more stable and reliable in operation. A likeness to the circuit of a monostable multivibrator may be noticed. We have similar cross-

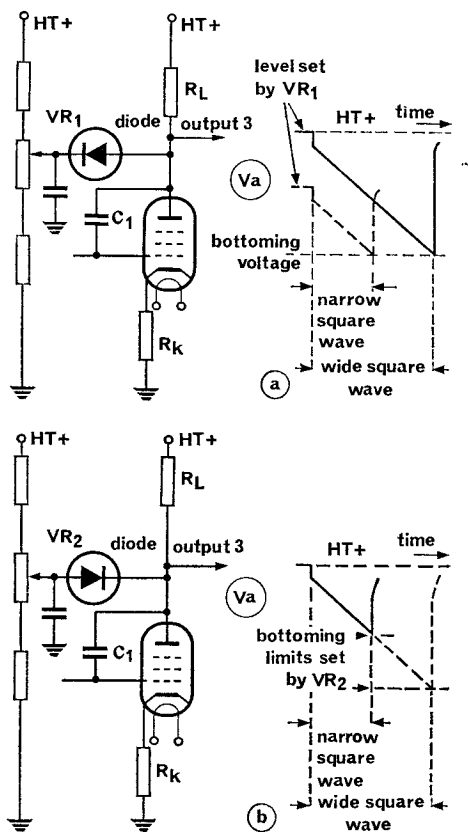


FIG. 4. VARIATION OF PULSE DURATION IN A PHANASTRON

coupling impedances between the two stages, one a resistor and the other a capacitor. Unlike the multivibrator however the connection from V_1 anode is to V_2 suppressor grid and V_2 also has a 'Miller' capacitor C_3 .

The arrangement and the action of V_2 are similar to those of the phantastron. V_1 provides the switching action which returns the circuit to its stable state at the end of the Miller rundown. The circuit may be triggered by a *negative* pulse either to the control grid or to the suppressor grid of V_1 . Suppressor grid triggering is more usual.

In the stable condition, before the application of a trigger pulse, V_1 is conducting normally with its control grid and suppressor grid both at zero volts. V_1 anode voltage is therefore *low* and is sufficient to keep V_2 suppressor grid below the *anode current cut-off*

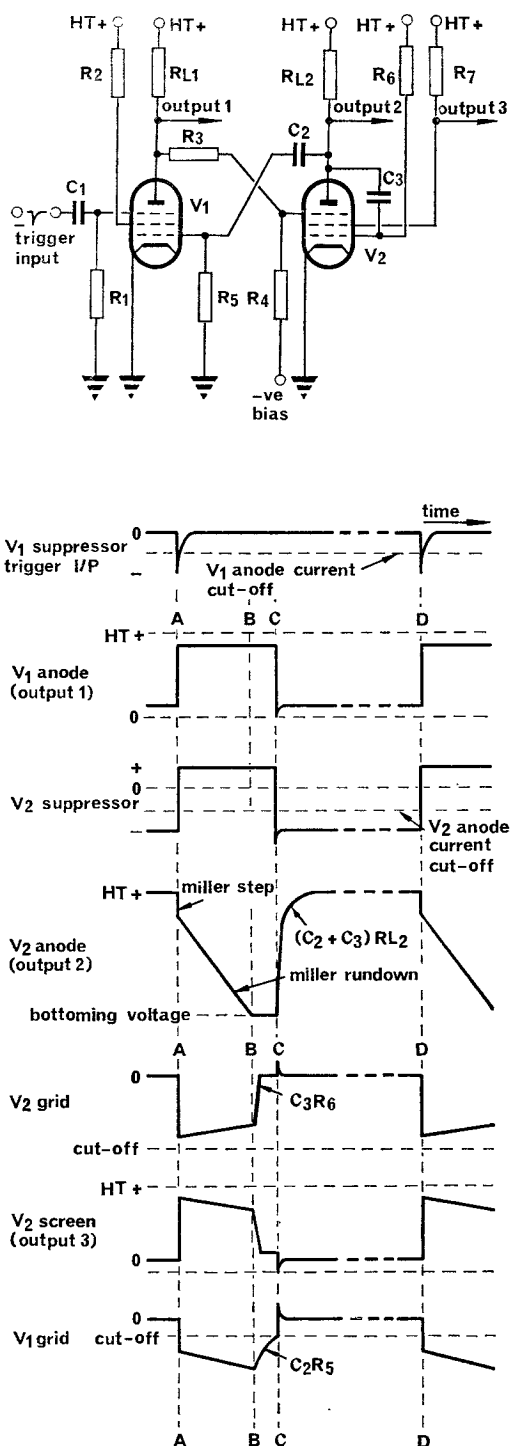


FIG. 5. SANATRON, CIRCUIT AND WAVEFORMS

point due to the voltage division in the potentiometer chain R_{L1} , R_3 , R_4 connected between h. t. + and the bias. The grid of V_2 is held at cathode potential by grid current limiting through R_6 so that V_2 is also conducting. V_2 anode current is, however, cut off by the voltage on the suppressor grid so that the total space current of V_2 flows to the *screen*. V_2 anode voltage is thus at h.t. + and its screen voltage is at a low value.

Instant A. The negative trigger input pulse cuts off anode current in V_1 and V_1 anode voltage rises, taking up the suppressor grid voltage of V_2 . *Anode* current now flows in V_2 and V_2 anode voltage *falls*, taking V_2 grid voltage with it (via C_3). This produces the normal 'Miller step' action in V_2 as explained for the phantastron. With the flow of *anode* current in V_2 , the screen current falls and V_2 screen voltage *rises*. The fall in V_2 anode voltage is also applied via C_2 to V_1 grid. V_1 grid base is shorter than that of V_2 so that the fall in V_2 anode voltage is sufficient to cut off V_1 at the grid and V_1 anode voltage remains at a high value.

Interval A to B. The usual Miller rundown now occurs, during which V_2 grid voltage rises slowly, the discharge of C_3 through R_6 being opposed by Miller feedback. V_2 anode voltage thus falls *linearly*. V_1 grid voltage is subject to two opposing forces: C_2 is trying to discharge through R_5 to *lift* V_1 grid voltage to zero, but at the same time V_2 anode voltage is *falling*, tending to take V_1 grid voltage with it. The circuit component values are made such that the fall in V_2 anode voltage has the greater effect so that V_1 grid voltage falls slowly. V_1 therefore remains cut off, V_1 anode voltage remains at a high value and V_2 suppressor grid voltage remains above the anode current cut-off level. V_2 screen grid voltage therefore remains at a high value.

Interval B to C. At B, V_2 anode voltage reaches its bottoming value and, with no Miller feedback, V_2 grid voltage rises rapidly to zero as C_3 discharges through R_6 . V_2 anode voltage *remains* at its bottoming voltage however, since V_1 is still cut off and V_2 suppressor grid is thus still above its anode current cut-off level. Because V_2 anode voltage has now stopped falling, V_1 grid voltage can rise to zero as C_2 discharges through R_5 . V_1 cuts on at point C. V_2 grid voltage rises more rapidly than that of V_1 because it is aiming for h.t. + whereas V_1 grid voltage is aiming for zero volts through a high value resistor (R_5).

Instant C. V_1 grid reaches cut-off and V_1 conducts. V_1 anode voltage therefore falls taking V_2 suppressor grid voltage with it. V_2 anode current therefore falls and its anode voltage rises making V_1 grid voltage rise further (via C_2). An avalanche therefore occurs which rapidly switches the circuit back to its initial stable state, where it remains until another trigger pulse is applied (at instant D).

A square wave output is obtained from the screen of V_2 and also from the anode of V_1 . V_2 anode provides a sawtooth timebase waveform. The pulse duration can be altered by varying the Miller rundown time with the aid of anode-catching diodes, as in the phantastron.

CHAPTER 9

RINGING AND BLOCKING OSCILLATORS

Introduction

In Section 1 (p 5) we saw how the range of a radar target is found by measuring on a c.r.t. the time interval between transmission and reception of a pulse of r.f. energy. The spot begins to move on the c.r.t. screen at the instant the transmitter fires each pulse and moves over a definite path to form a trace. The spot must move at a *constant speed* and reach the end of the trace after a time determined by the maximum required radar range. Reflected pulses are amplified by the receiver and used either to deflect or brighten the trace at the instant they are received. A simple type A display is illustrated in Fig 1. To ensure accuracy of range measurement we must check that the spot moves at the correct speed. *Calibration markers* (commonly called 'cal pips' or 'range markers') can be provided to assist in this.

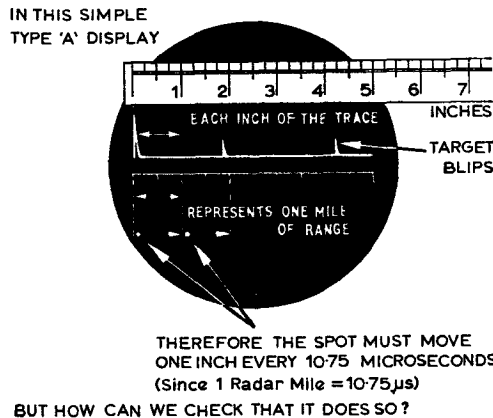


FIG 1. ACCURACY OF RANGE MEASUREMENT IN A TYPE 'A' DISPLAY

Cal Pips

To calibrate any device means to 'check and correct for any irregularities'. Thus when we check to ensure that the spot is moving along the trace at the correct speed we are *calibrating* the display.

One method of calibration is to generate a series of very narrow pulses at intervals of exactly one radar mile (10.75μ s), and apply them to the c.r.t. so that they deflect or brighten the trace to give cal pips at the correct intervals along the trace. If the spot is moving at the correct speed

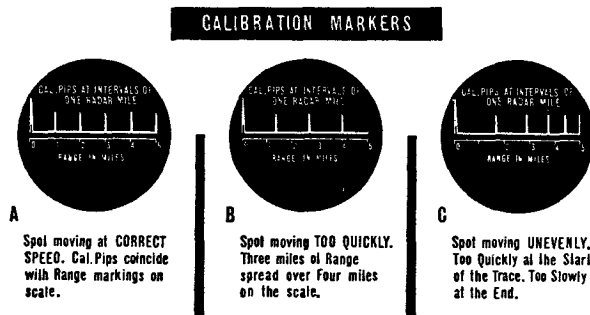


FIG 2. USE OF CAL PIPS FOR RANGE CALIBRATION, TYPE 'A' DISPLAY

(Fig 2a) the cal pips line up exactly with the mile markings on the fixed range scale. If they do not line up (Fig 2b and c) the circuits which control the spot movement (i.e. the timebase circuits) need to be adjusted.

In Fig 2 the maximum range of the equipment is shown as only five miles and cal pips at intervals of one mile range are convenient for calibrating the display. If the maximum range is 100 miles, 100 cal pips along a few inches of trace would be confusing; it is then more convenient to have markers at, say, 10-mile intervals.

Many radar equipments have more than one range. Thus when switching from one range to another the time intervals between cal pips must also be changed to avoid cluttering up the display with too many cal pips at the longer ranges.

In the type A display the cal pips deflect the trace to give blips. In intensity-modulated displays the cal pips are applied to the control grid or the cathode of the c.r.t. to produce bright spots on the screen. In a p.p.i. display, where the trace is rotating, the bright spots form *calibration rings* on the face of the c.r.t. (Fig 3).

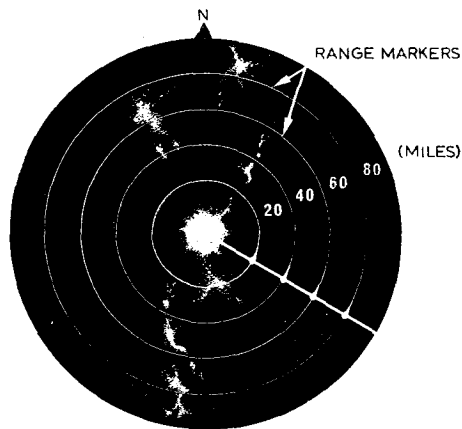


FIG 3. PPI RANGE MARKERS

Cal pips can also be used to measure the *pulse duration* of a square wave (Fig 4). The waveform being examined and the cal pips are applied together to an oscilloscope. The period and pulse duration of the waveform are then found simply by counting the cal pips. Knowing the time interval between successive cal pips, the other factors follow.

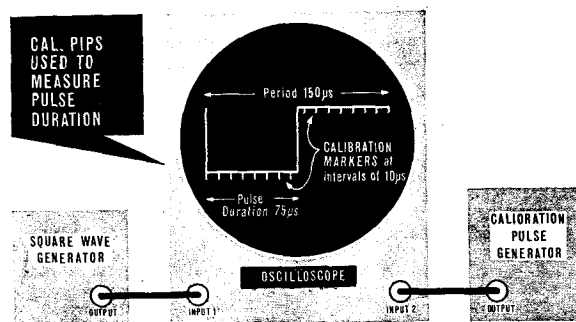


FIG 4. CAL PIPS USED TO MEASURE PULSE DURATION

We have now seen some examples of how cal pips are used. In each case the circuit which generates the cal pips must produce very narrow pulses of voltage, and the time interval between successive pulses must be very accurately known. With these requirements in mind let us now see how cal pips are produced.

Production of Cal Pips from Sine Waves

Very accurate cal pips can be produced by the arrangement shown in block form in Fig 5. The arrangement is known as a *cal pip generator*. The oscillator generates a sine wave at an accurate frequency. The sine wave is then squared, differentiated and applied to a limiter, where it is positively limited from a fixed negative voltage. Finally, the narrow negative-going pips from the limiter are applied to a pulse shaper which gives the output waveform shown in Fig 5; this is a series of narrow positive-going pulses of exactly the same frequency as that of the oscillator.

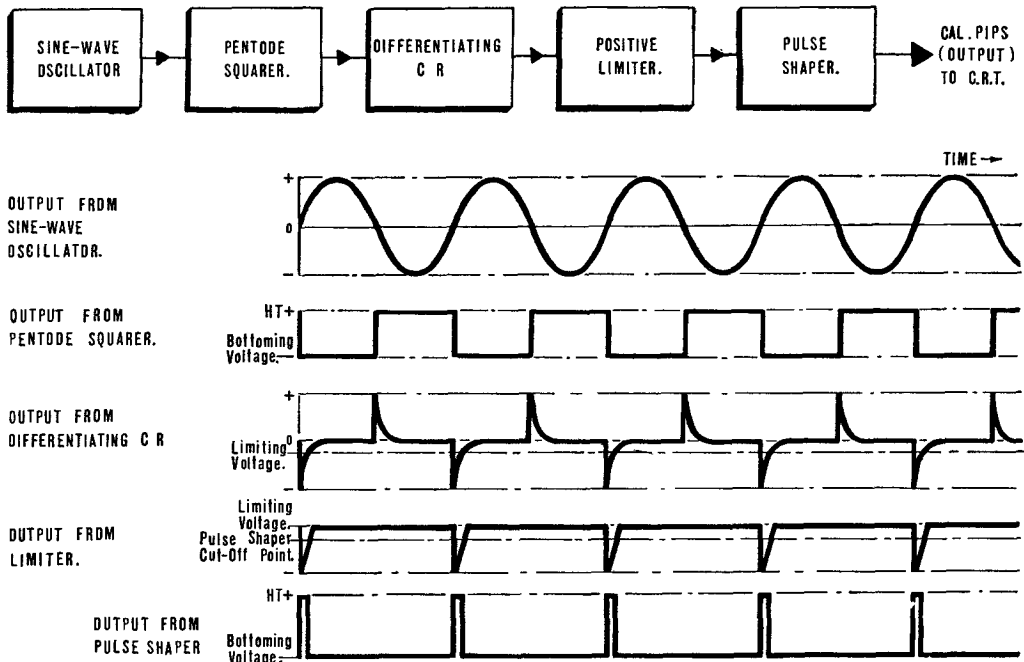


FIG 5. CAL PIP GENERATOR AND WAVEFORMS

If we require cal pips at intervals of one radar mile, the period of the output waveform must be $10.75 \mu\text{s}$ ($\frac{1}{93,000}$ second). The frequency of the oscillator must therefore be 93 kc/s.

Cap Pip Generator Switching

When we use cal pips on a radar display as range markers, the first pip on each trace must occur at the instant the spot begins to move, otherwise the other pips will not be correctly positioned along the trace. We must therefore *trigger* the cal pip generator in some way to ensure correct synchronization between the trace and the cal pips.

We normally use a pulse generator, such as a flip-flop, to produce square waves which switch on the indicator circuits at the instant the transmitter fires and switch them off when echoes have returned from maximum range. This same switching waveform is used to switch on the cal pip

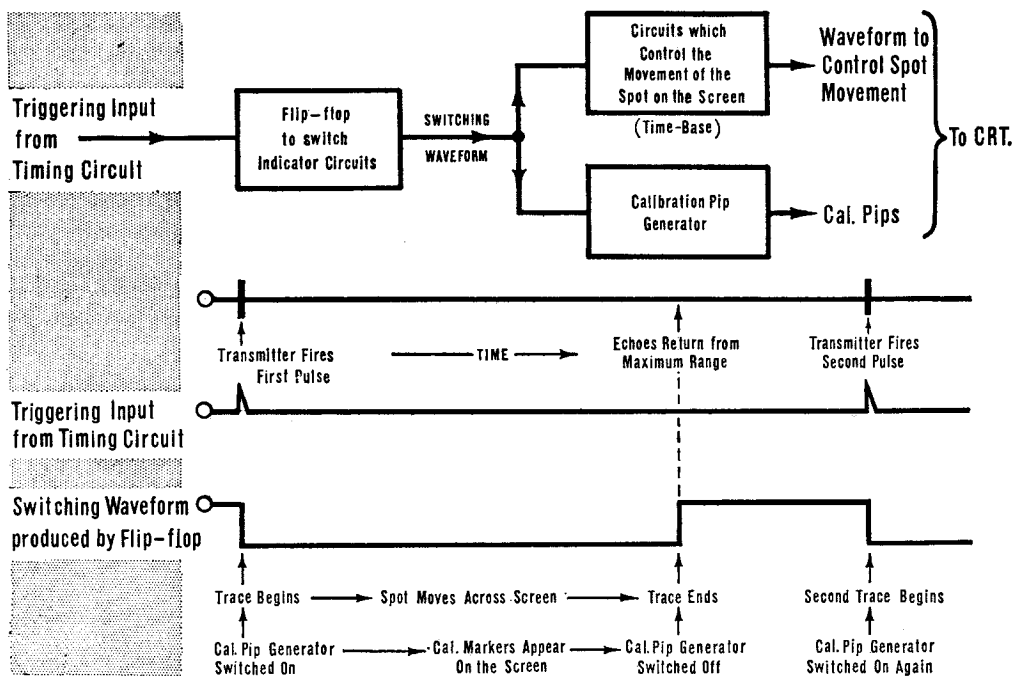


FIG 6. CAL PIP GENERATOR SWITCHING

generator at the instant the trace begins and switch it off when the trace ends. The arrangement is shown in Fig 6.

Ringng Oscillator

The oscillator most commonly used in a cal pip generator is the *ringng oscillator*, the circuit and waveforms of which are shown in Fig 7. As explained in the previous paragraph, the oscillator is being alternately switched on and off by the same square wave input that is being used to trigger the timebase circuits. When the valve is suddenly switched on at A, current flows into the tuned circuit, making it ring (see p 81). The conducting valve, however, is effectively in parallel with the tuned circuit and very quickly *damps* the oscillations. At B when the valve is cut off the supply current to the tuned circuit ceases abruptly, the magnetic field around the inductor collapses, inducing a back e.m.f. into the circuit, and the tuned circuit again rings. This time however, with the valve cut off, there is very little damping of the tuned circuit and the oscillations decay slowly. The frequency of oscillations is determined by the resonant frequency of the tuned circuit and this may be varied by adjustment of the pre-set capacitor C_2 . This, in turn, varies the time interval between successive cal pips. For five-mile cal pips, the interval between each pip is $5 \times 10.75 \mu s = 53.75 \mu s = \frac{1}{18,600}$ second, and the frequency must be adjusted to 18.6 kc/s.

When the resonant circuit is placed in the anode of the valve, as in Fig 7, the first half-cycle of ringing, when the valve is cut off, is *positive-going*. Under certain circumstances the first half-cycle of ringing is required to be *negative-going*. This is obtained by placing the tuned circuit in the *cathode* of the valve.

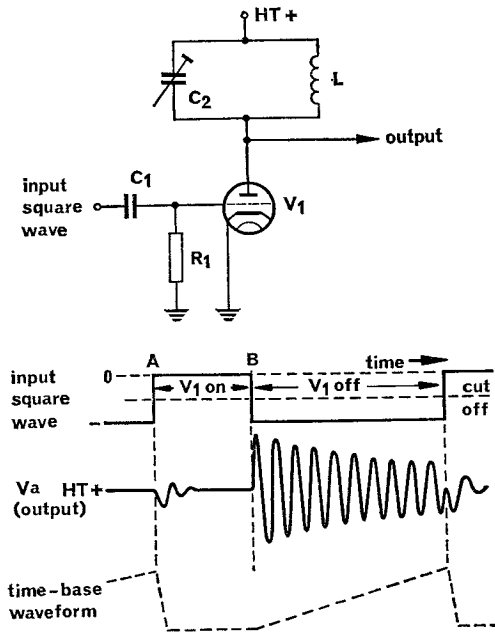


FIG 7. RINGING OSCILLATOR, CIRCUIT AND WAVEFORMS

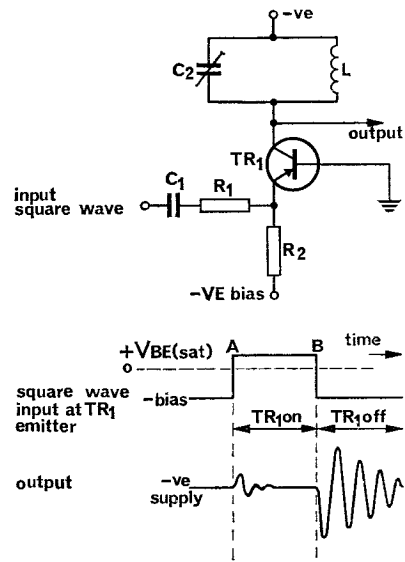


FIG 8. TRANSISTOR RINGING OSCILLATOR

The ringing oscillator needs to be checked frequently to ensure that its operating frequency is correct. In most radar systems a *crystal-controlled* oscillator is supplied for use as a check on the frequency. The crystal-controlled oscillator cannot itself be used in place of the ringing oscillator, because of difficulties in triggering.

Transistor Ringing Oscillator

Like most circuits, we may also have a transistor version of the ringing oscillator. A common-base arrangement of a ringing oscillator, together with associated waveforms, is shown in Fig 8. For a p-n-p transistor a negative-going input to the *base* or a positive-going input to the *emitter* is required to cut the transistor on. In Fig 8, where the input is applied to the emitter, when the square wave rises at A the transistor cuts on and the circuit rings. This train of oscillations is quickly damped by the conducting transistor. When the square wave input cuts the transistor off at B the circuit again rings—this time with very little damping. The resulting train of oscillations may be used to produce cal pips, as explained earlier.

Ringling Oscillator as Pulse Generator

The ringing oscillator may be modified to produce a series of *single* pulses (instead of a train of pulses), each pulse being tied to the *trailing* edge of the input square wave. The required modifications to the circuit are indicated in Fig 9a:

- The inductor L may contain a variable core to provide adjustment of frequency.
- The capacitor C is reduced to a very small value, very often only the self-capacitance of the coil.
- A resistor R may be inserted across the coil.
- A diode D may be connected across the coil in the direction indicated.

The oscillator output with only L and C connected is as previously described and is indicated in Fig 9b.

If we now insert the resistor R , of such a value that the circuit is *critically damped*, the output is as shown in Fig 9c.

If we now insert the diode D , in addition to R , the diode limits the *negative* swings of anode voltage and the output is as shown in Fig 9d.

Note that if D is inserted and R *removed* the circuit is no longer critically damped so that a large-amplitude positive-going swing can be obtained. The diode conducts on the negative-going portions of the output waveform as before so that only a large positive-going pulse, coincident in time with the trailing edge of the input square wave, is obtained (Fig 9e).

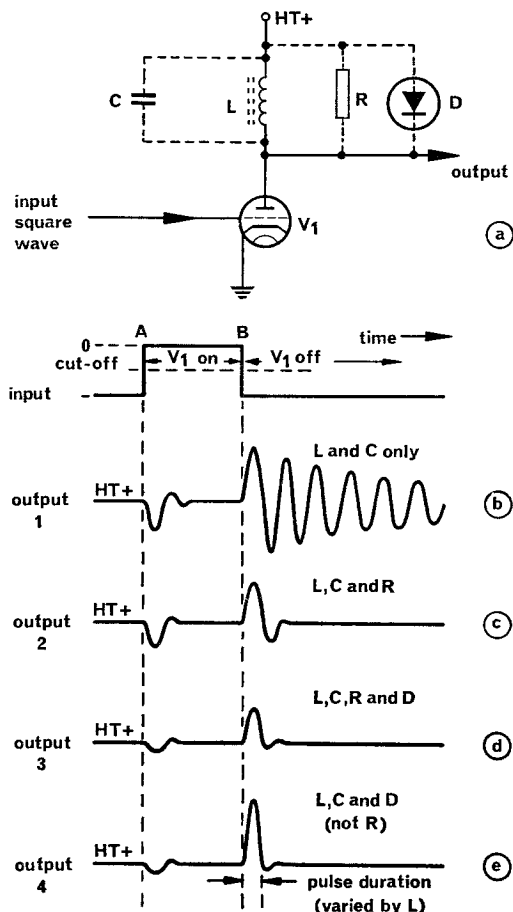


FIG 9. RINGING OSCILLATOR PRODUCING SINGLE PULSES

Normally, *both* R and D are inserted to prevent an excessive positive voltage swing at the anode and to ensure a greater control over the pulse duration of the output waveform.

The output from this type of ringing oscillator may be used as the trigger pulse for the modulator in a radar transmitter. If it is used as such the transmitter pulse duration depends upon the duration of the output pulse from the ringing oscillator. This, in turn, may be varied by adjusting the tuning slug of the inductor L .

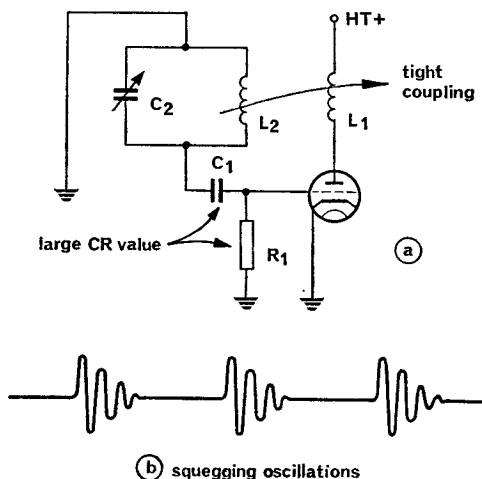


FIG 10. TUNED GRID OSCILLATOR UNDER SQUEGGING CONDITIONS

Blocking Oscillator

The blocking oscillator produces a series of very narrow, steep-sided pulses of large amplitude which can be used as cal pips or, more usually, as trigger pulses. The waveform at the control grid has also been used in some television receivers to provide a sawtooth timebase waveform.

The blocking oscillator is, in effect, a development of the tuned anode or tuned grid oscillator considered in Part 1B of these notes (p 352). In the normal tuned grid oscillator (Fig 10a), if the feedback between anode and grid is great enough, and the CR time constant of the bias

circuit long enough, the charge which builds up on C_1 when grid current flows during the positive-going half-cycles of grid voltage is sufficient to bias the valve well beyond cut-off. Oscillations then cease until C_1 has discharged sufficiently to cut the valve back on. As a result, the circuit oscillates in 'bursts' (Fig 10b). An oscillator working under these conditions is termed a '*self-quenching*' or '*squegging*' oscillator.

In the blocking oscillator the inductive coupling between anode and grid is so 'tight' and the CR value so large that squegging is *deliberately* introduced, to such an extent that the valve 'blocks off' very rapidly. The output is only a single pulse of voltage.

The blocking oscillator is a member of the *relaxation oscillator* family (like the multivibrator). In any relaxation oscillator, the valve conducts heavily for a given period of time and then cuts off, this being followed by a relaxation period during which the valve is returning to the state where it will again conduct.

The basic circuit of a blocking oscillator, with its associated waveforms, is shown in Fig 11. The tight coupling between anode and grid is provided by an iron-cored pulse transformer. The tuned circuit capacitance is usually omitted, the self-capacitance of the windings being sufficient.

If we assume that V_g is rising towards cut-off, then at instant A the valve conducts, the anode current rises and V_a falls. The rising current in L_1 induces a voltage in L_2 , the direction of the windings being such that V_g rises. The anode current therefore rises further, V_a falls further and an avalanche occurs. V_g is driven *above zero* volts and V_a falls to a low value.

When V_g rises above zero, grid current starts to flow, charging C_1 in such a direction as to *tend* to make V_g *negative*. The voltage developed across C_1 *opposes* the e.m.f. induced across L_2 and the feedback is then insufficient to maintain the increasing anode current. At instant B, the anode current begins to fall causing V_g to *fall*. The anode current therefore falls further and another avalanche occurs in the reverse direction. V_g is now driven well below cut-off and V_a rises to h.t. +.

With the abrupt cessation both of anode and grid currents, *ringing* occurs in the anode and grid circuits and damped oscillations at the ringing frequency are produced. The damping in the grid circuit is made high enough to prevent the positive-going portion of the grid ring rising above the cut-off level.

During the interval B to C the capacitor C_1 is discharging through R_1 towards zero volts. When V_g rises to cut-off, the valve again conducts and the action is repeated as from instant A.

In this circuit the output is taken from the anode. It consists of a series of *negative-going*, large-amplitude pulses with very steep edges (the

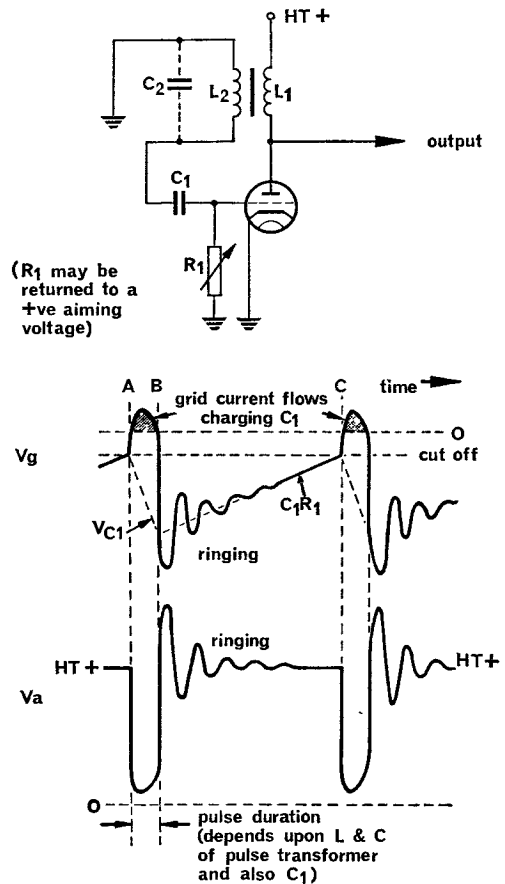


FIG 11. BLOCKING OSCILLATOR, CIRCUIT AND WAVEFORMS

leading and trailing edges are both the result of avalanches). Very short pulse durations, of the order of $1\mu\text{s}$, are possible. The pulse duration is determined by the inductance and self-capacitance of the pulse transformer and also by the value of C_1 ; for a short pulse duration, C_1 should be small. The p.r.f. of the circuit is determined by the time taken for V_g to return to cut-off. This may be controlled by varying R_1 or C_1 or by taking R_1 to a variable positive aiming voltage as in the multivibrator.

It is also possible, by placing a resistor R_2 in the cathode of the valve, to take the output from the *cathode*. The output pulses are now *positive-going*.

The blocking oscillator may be *synchronized* in much the same way as a multivibrator. If we apply a series of positive-going sync pulses to the grid such that V_g rises above the cut-off level at each sync pulse—before it would normally do so—then each output pulse is synchronized with the corresponding input pulse.

The ringing which occurs in the anode circuit may be prevented by connecting a diode across the output winding as shown in Fig 12. The diode is connected in such a way that it plays no part in the circuit action until V_a tries to rise above h.t. +; it then conducts to hold the output at h.t. +. As the first half-cycle of the ringing action cannot now take place the whole effect is eliminated and the output waveform is then as shown in Fig 12.

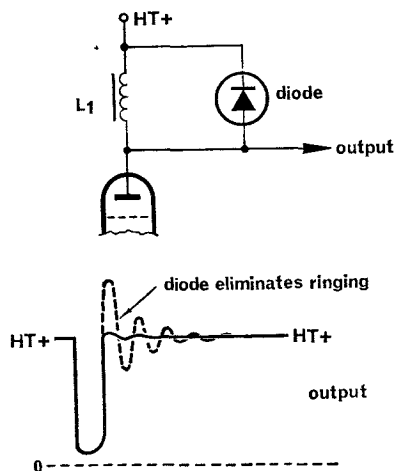


FIG 12. PREVENTION OF RINGING IN BLOCKING OSCILLATOR

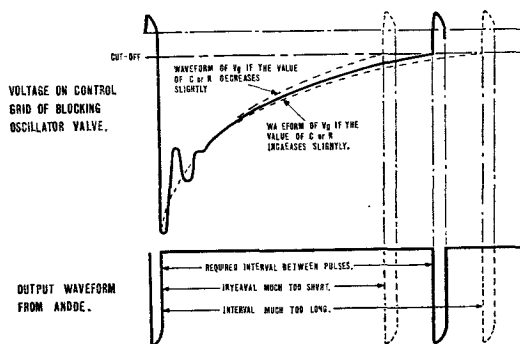


FIG 13. VARIATION OF PRF IN FREE-RUNNING BLOCKING OSCILLATOR

Triggered Blocking Oscillator

The astable, *ie* free-running, blocking oscillator considered in the previous paragraphs is sometimes used to produce the master timing pulses for a complete radar system, thus determining the instant of time at which each circuit in the radar operates. However, the p.r.f. of a free-running blocking oscillator is erratic due to slight variations in the circuit conditions (see Fig 13). Because of this it is usual to synchronize the operation of the blocking oscillator with pulses whose p.r.f. can be accurately controlled.

However, under other circumstances, the requirement is that the blocking oscillator produces no output *until it is triggered*, *ie* a monostable circuit may be required. The circuit and waveforms of a triggered blocking oscillator are shown in Fig 14. R_1 is now returned to a negative bias voltage which is sufficient to keep the valve cut off under normal conditions. To cut the valve on and produce *one cycle* of blocking oscillator operation the circuit must be triggered by applying a positive pulse to the grid, of sufficient amplitude to lift V_g above cut-off. As soon as V_g rises above cut-off the action is as described for the free-running circuit.

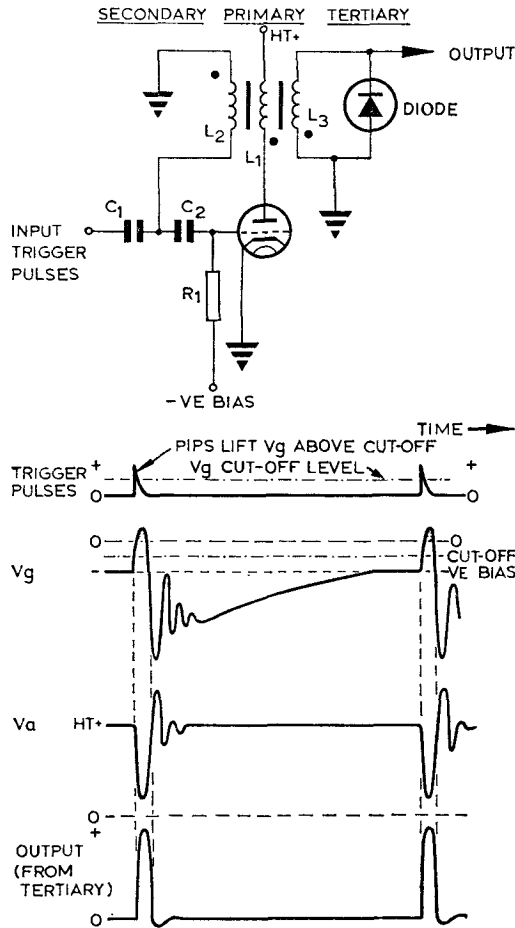


FIG 14. TRIGGERED BLOCKING OSCILLATOR, CIRCUIT AND WAVEFORMS

The circuit of Fig 14 contains a pulse transformer with *three* windings, the output being taken from the *tertiary*. This is often used when we require large-amplitude *positive-going* output pulses. The overswing diode in this case is connected to limit the *negative* overshoot caused by ringing.

Transistor Blocking Oscillator

The circuit of a transistor blocking oscillator consists essentially of a transistor with transformer-coupled positive feedback from the collector to either the base or the emitter, and with a RC network in either the emitter or the base circuit. The four possible circuit arrangements are shown in Fig 15. The action of these four circuits is essentially the same and is very similar to the action of the valve blocking oscillator described earlier.

Fig 16 shows a practical *triggered* blocking oscillator based on the arrangement given in Fig 15d. A grounded-base circuit is used, with both the feedback winding L_2 and the timing circuit C_2R_2 in the *emitter* circuit. Bias is applied to the emitter via R_1 , the values of R_1 and R_2 being such that the emitter is sufficiently *negative* with respect to its base to keep TR_1 cut off. To cut the transistor on, a *positive-going* trigger pulse is applied to TR_1 *emitter* via the coupling capacitor C_1 and the resistor R_3 (inserted for matching). A diode and a damping resistor R_4 are connected across the transformer primary L_1 to prevent excessive ringing at the collector.

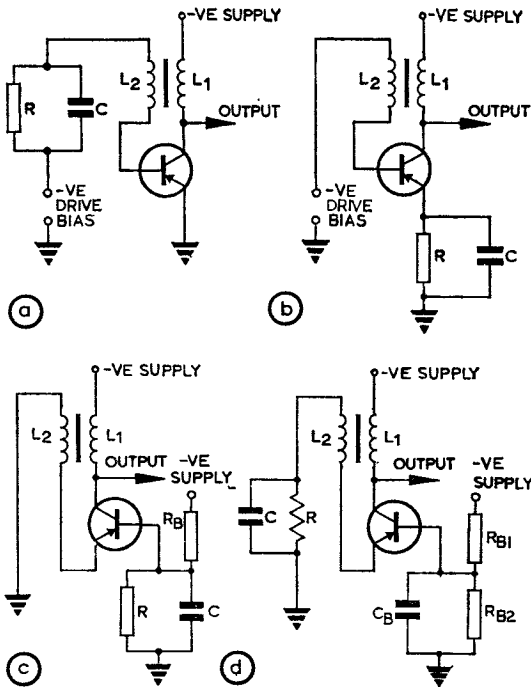


FIG 15. TRANSISTOR VERSION OF BLOCKING OSCILLATOR

By choosing suitable values for the components in this circuit, very short pulse durations of the order of $1 \mu s$ at a p.r.f. of 100 kc/s can be obtained, ie 100,000 one-microsecond pulses every second!

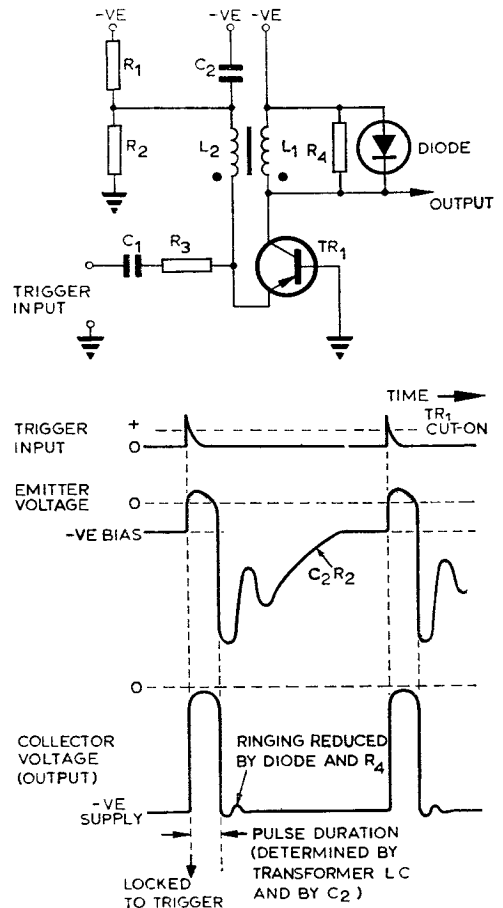


FIG 16. TRIGGERED BLOCKING OSCILLATOR, GROUNDED-BASE CIRCUIT

Unijunction Transistor as Pulse Generator

The unijunction (single junction) transistor consists of a bar of n-type semiconductor with two non-rectifying contacts (base 1 and base 2), one at each end, as shown in Fig 17a. A third connection (rectifying this time) is made to the other side of the bar near base 1. If bases 1 and 2 are connected together externally, the device behaves as a normal p-n junction diode. However under normal conditions a voltage V_B is applied between base 1 and base 2, such that base 2 is biased positively with respect to base 1. A very small inter-base current therefore flows. With no emitter current flowing, the bar behaves as a potential divider with a fraction X of V_B appearing at the emitter junction. This voltage effectively *reverse-biases* the emitter junction and little current flows in the device. However if we apply a *positive-going* voltage to the emitter *greater* than XV_B the emitter junction is now *forward-biased* and a large emitter-base 1 current flows.

The unijunction transistor may therefore be used as a pulse generator by connecting it in the circuit shown in Fig 17b. The action is simple: the capacitor C_1 charges through R_1 towards the supply voltage level $+V_B$. When the capacitor voltage reaches a value greater than $+XV_B$ the unijunction 'fires' and C_1 quickly discharges into the emitter circuit. The voltage developed across R_2 by this discharge current provides the output pulse. When the emitter voltage has

fallen to a low value the emitter junction again becomes reverse-biased and the unijunction switches off. The process is then repeated at a p.r.f. determined by the charging time constant $C_1 R_1$ seconds.

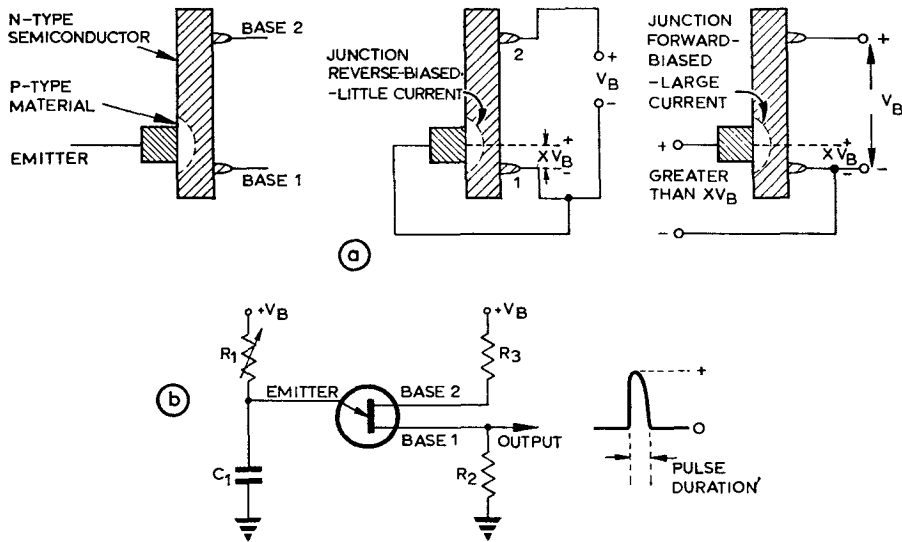


FIG 17. THE UNIJUNCTION TRANSISTOR, USE AS A PULSE GENERATOR

CHAPTER 10

ELECTRONIC SWITCHING CIRCUITS

Introduction

We have already seen that in a pulsed radar equipment many circuits have to be switched on and off very rapidly. Very often one circuit has to be brought into operation at *exactly* the same instant of time as some function occurs in another circuit. Alternatively, we may require a circuit to become operative an *exact* number of microseconds *after* some function has occurred elsewhere in the equipment. Because of the very short time intervals involved—often of the order of *nanoseconds* (i.e. 10^{-9} seconds)—mechanical switches and relays cannot be used. Electronic switches are therefore necessary.

Switching is carried out at various stages throughout a radar equipment. In the initial stages of a trigger unit (e.g. at the master timing unit) the switching is very often at millivolt and microampere levels. At intermediate stages, switching up to several hundreds of volts at a few milliamperes may be required. At the final transmitter stages the power levels are very high and switches capable of handling several thousands of volts at currents of the order of tens of amperes are necessary. The final switching stage of one ground radar transmitter switches at a voltage of 38,000V and at a current of about 130 amperes.

Electronic Switches

The earliest electronic switch was the thermionic valve. A valve is easily cut on and off, and the speed at which the action occurs is high. Fig 1b illustrates the switching action. Hard valve switches have certain limitations, among them the fact that the power they can readily handle is limited. Where fast switches capable of passing very high currents at high voltages are required a gas-filled valve (such as a thyratron) may be used.

A transistor may also be used as a switch in much the same way as a valve by applying suitable voltages to the base-emitter circuit (Fig 1c). A semiconductor diode is another good electronic switch (Fig 1d).

Tunnel Diode as a Switch

All the devices mentioned are limited in their switching speed. In valves, transit time is the main limitation and in semiconductor devices it is the hole-storage effect (see Part 1B, p 322). Where very high switching speeds at very low power levels are required a *tunnel diode* may be used. This is a semiconductor device capable of nearly instantaneous switching. With the tunnel diode, switching times of less than one nanosecond have been achieved.

The tunnel diode looks like a normal p-n junction diode. But the p and n materials are more heavily 'doped', having a larger number of acceptor and donor impurities, and the junction

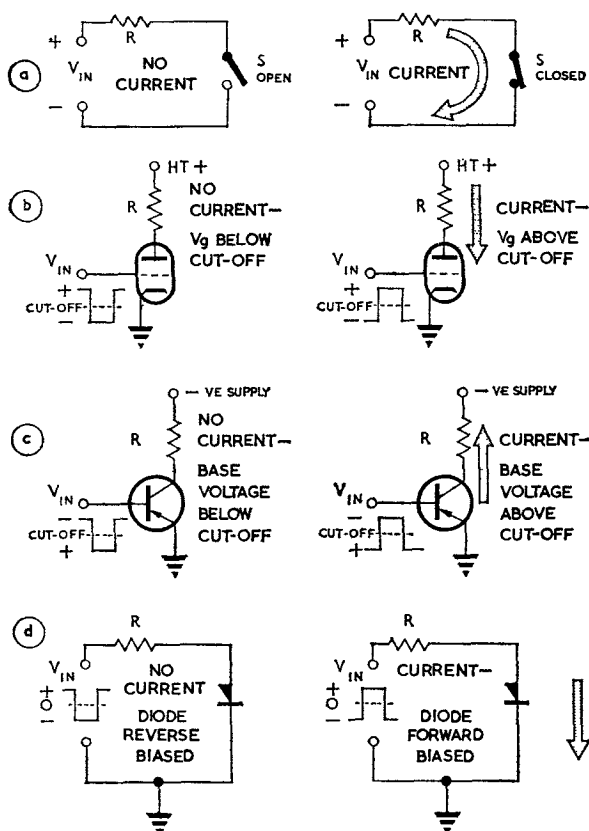


FIG 1. ELECTRONIC SWITCHES

is extremely thin. Because of this, electrons can 'tunnel' through the very thin junction at low values of forward-bias voltage. As the forward bias is increased the tunnel current increases to a peak value at A (Fig 2). It then *falls* with further increase in forward bias to a level known as the 'valley current' at B. Thereafter the characteristic is that of a normal forward-biased junction diode. For a germanium tunnel diode the peak and valley current points occur at about 50mV and 350mV respectively.

Notice that between points A and B the current is *falling* as the voltage is *increased*. The tunnel diode therefore behaves as a *negative resistance* over this region. By suitable arrangements, such a device can be made to provide *amplification*. A tuned circuit load has a certain dynamic impedance at its resonant frequency. If a negative resistance is placed *in parallel* with this load, the normal (positive) resistive damping losses of the tuned circuit may be completely cancelled by the negative resistance and the circuit will then *oscillate*. By suitable adjustment of the standing bias on the tunnel diode the negative resistance value is made such that the circuit does *not quite* oscillate. The arrangement now works as a tuned amplifier. This is something quite new for a diode.

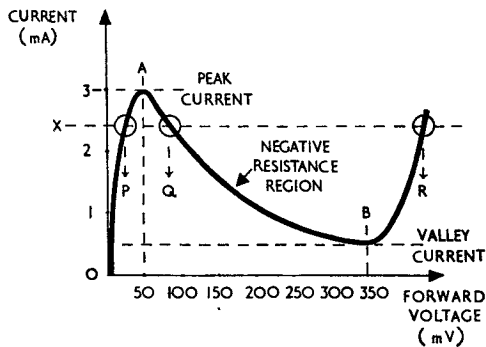


FIG 2. TUNNEL DIODE CHARACTERISTIC

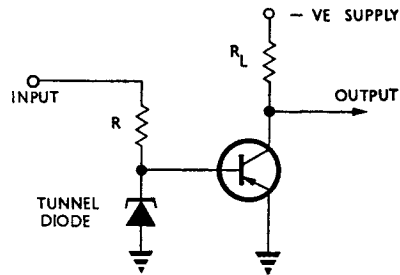


FIG 3. TUNNEL DIODE-TRANSISTOR TRIGGER CIRCUIT

The two main applications of the tunnel diode are in switching circuits, where the device is switched from the peak to the valley points by an input current, and in low-noise amplifiers and oscillators for extremely high frequencies, where the tunnel diode is biased in its negative resistance region.

In the switching application, the characteristic of Fig 2 shows that a given value of current x may correspond to three different applied voltage levels, P, Q and R. In most switching circuits, voltages between A and B are unstable and the tunnel diode is usually considered as a *two-state* (bistable) device, one state to the left of the peak current point and the other to the right of the valley point. To switch from one state to the other a small change in input current in one sense or the other is all that is required.

A useful tunnel diode-transistor trigger circuit is shown in Fig 3. When the circuit is first switched on it is arranged that the tunnel diode is switched to the peak current-low voltage region. The base emitter junction of the transistor has therefore a low voltage across it and the collector current is negligible. The output voltage is thus at the negative supply voltage level. As the input current is increased the tunnel diode switches rapidly to the valley current-high voltage state and the transistor conducts heavily. The output voltage thus rises rapidly towards earth. By changing the input current by small amounts in this way very rapid changes in output voltage may be obtained. High-speed low-voltage switching circuits of this type find many uses in computers.

Quench Circuit as a Low-power Switch

Fig 4 shows how an oscillator may be controlled by means of a 'quench' circuit. TR_1 is the quench or switching transistor which is normally conducting, with its base held at zero volts by the flow of base current through R . TR_2 and its associated circuit form an inverted Hartley oscillator. Since TR_1 is normally conducting, the current drawn by it through the oscillator coil L effectively damps the oscillator tuned circuit and prevents oscillations. Only when TR_1 is switched off by the positive-going trigger pulses on its base is the oscillator free to oscillate for the duration of the trigger pulse. This low-power circuit may be used as the first stage in a 'cal pip generator' (see p 145).

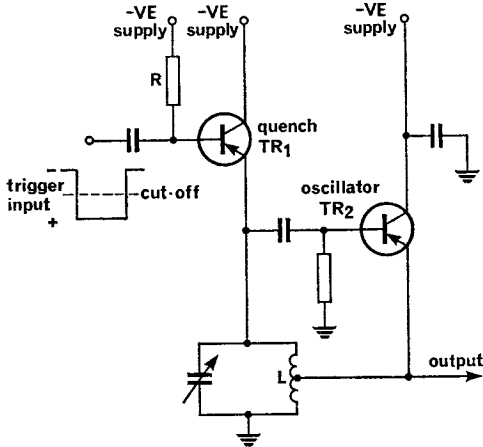


FIG 4. USE OF QUENCH CIRCUIT AS A SWITCH

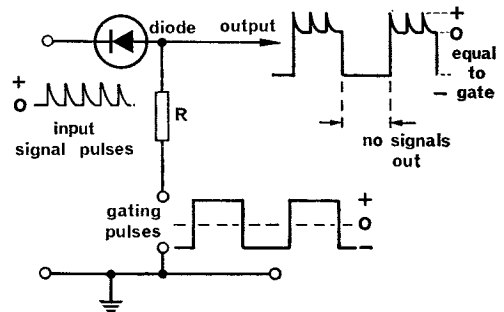


FIG 5. BASIC DIODE GATE

Gating Principles and Circuits

Gating is a switching process which may be likened to opening a gate to allow the passage of something only at and for a required time. In radar, gating circuits are used to allow input signals to pass to the output terminals at definite instants of time and for required periods of time. Gating is normally carried out at low-voltage, low-power stages in an equipment.

A simple diode gate is illustrated in Fig 5. The input consists of a train of positive-going pulses. A *gating* square wave is applied to the diode as shown. On the positive-going half-cycles of the gating waveform the diode conducts and the input signal pulses then appear at the output terminals. On the negative-going half-cycles of the gating waveform the diode is cut off and there is no output—even although the signal pulses are still present at the input.

Fig 6 illustrates a *pentode gate* together with its associated waveforms. The valve is normally biased beyond cut-off both on the control grid g_1 and on the suppressor grid g_3 . The gating square wave is applied to one grid (g_3 in this example) and the signal input is applied to the other grid g_1 . The gating square wave lifts the suppressor grid above cut-off during the periods AB and CD and the signal input to g_1 then causes the valve to conduct. Thus the only signals which produce an output are those which appear at the control grid g_1 during the positive-going portions of the gating waveform, when anode current is allowed to flow. The circuit thus selects input signals according to the time they arrive. A gating circuit is therefore a *time selection* circuit. The diode D is normally inserted to limit the suppressor grid voltage to zero volts.

A simple *transistor gate* is illustrated in Fig 7. The base-emitter junction of the transistor is normally held cut off by the negative bias applied to the emitter. The positive-going signal input pulses are applied to the *emitter* but these are of insufficient amplitude to cut on the

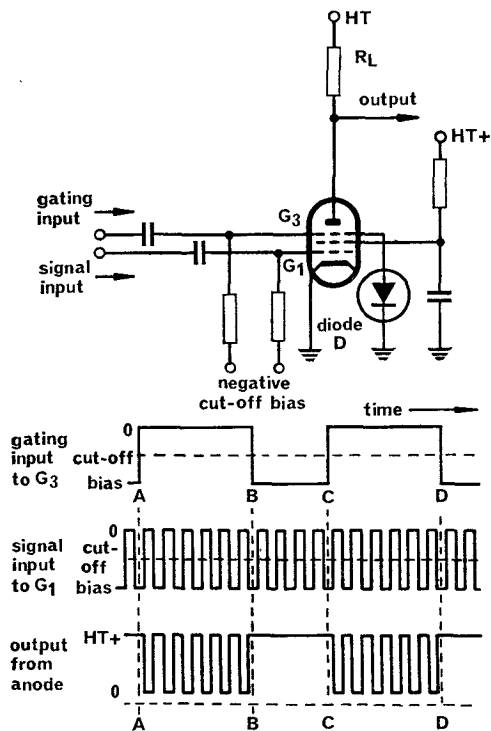


FIG 6. PENTODE GATE

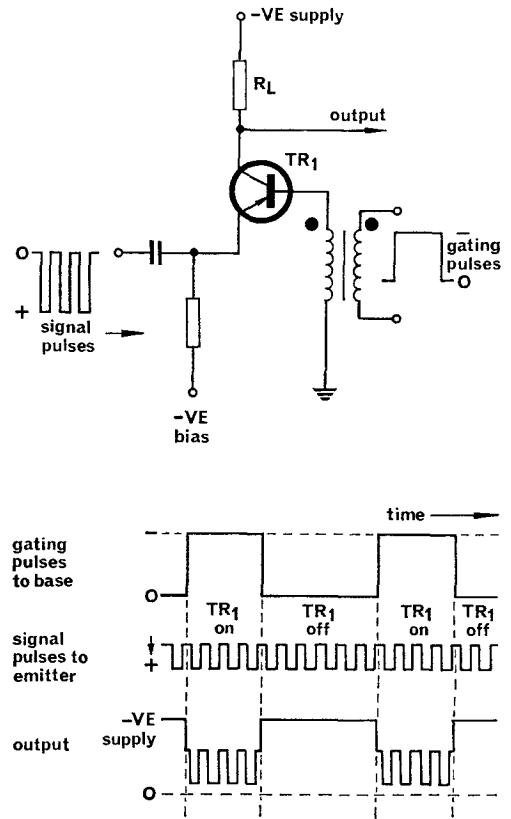


FIG 7. TRANSISTOR GATE

transistor. The transistor is cut on by *negative-going* gating pulses injected into the base via the transformer. The transistor therefore conducts for the periods of the negative-going portions of the gating pulses and during these times the signal pulses appear at the collector.

Coincidence Circuits.

A coincidence circuit and a gating circuit are very similar and so also is their action. The difference lies in the inputs applied to the two systems.

Fig 8a illustrates a pentode gate used as a coincidence circuit. From the waveforms of Fig 8b it is seen that an output is obtained only during the periods when *both* inputs are above cut-off, *ie* when the two waveforms *coincide*. In general, a *gating* circuit is used to provide an output only during required and definite intervals of time, *ie* it *selects* part of a signal with respect to time. A *coincidence* circuit, on the other hand, provides an output only when two separate functions occur *simultaneously*. We shall see an example of the use of coincidence circuits in p 214.

Trigger Circuits

We have already seen that many circuits, including the monostable and bistable multi-vibrator, operate only when a trigger pulse input is available. Trigger pulses are obtained in many ways and we have seen examples in earlier chapters. A square wave may be differentiated and then limited to produce either positive- or negative-going pulses. Where larger amplitude

trigger pulses of controlled pulse duration are required, the output from a critically-damped ringing oscillator or from a blocking oscillator or unijunction circuit may be used. We have also seen earlier in this chapter how a tunnel diode and a transistor may be used together to produce a trigger pulse output. Most of these circuits operate at low-voltage, low-power levels.

Most radar equipments contain a 'trigger unit'. This normally consists of a series of circuits so arranged that the initial trigger pulse is increased in amplitude and given the correct shape and time duration to trigger subsequent stages at precise instants of time.

The output from the trigger unit is of sufficient amplitude to trigger the final switching stage which, in turn, is used to turn the transmitting valve (e.g. a magnetron) on and off. The stage which controls the transmitting valve in this way is referred to as the *modulator*.

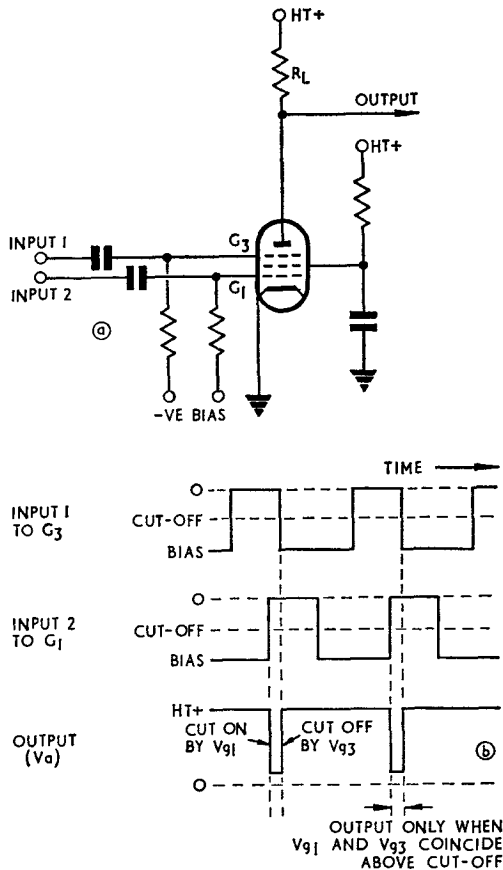


FIG 8. COINCIDENCE CIRCUIT

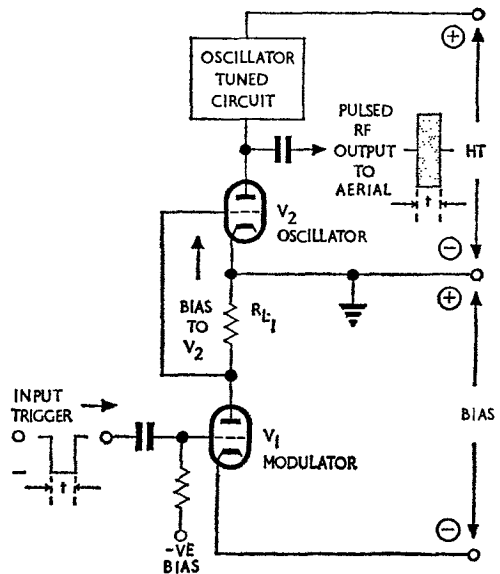


FIG 9. BASIS OF GRID MODULATION IN A PULSED RADAR TRANSMITTER

Modulators

The power requirements of the modulator depend to a large extent upon the type of transmitting valve used. We shall consider this in more detail in Section 5. It is sufficient at this stage to know:

- If the transmitting valve is a magnetron, which must be anode-modulated, the modulator must be capable of handling the full peak pulse power. This can be of the order of megawatts.
- If a klystron amplifier is used as the transmitting valve it can be switched on and off by power applied to a modulating electrode. In this case the modulator is required to handle only a small portion of the total output power.

c. If a disc-seal triode or other similar grid-controlled valve is used in the transmitter, grid modulation can be used. The modulator can now be a low-power valve.

Hard Valve Modulator

Fig 9 shows in outline how *grid modulation* of the final stage of a relatively low-power pulsed radar transmitter may be achieved.

V_2 is the transmitting valve and V_1 the modulator. In this circuit V_1 is normally conducting and passing a heavy current. The voltage developed across R_{L1} by V_1 current is applied as *bias* to the grid-cathode circuit of V_2 . Under normal conditions this bias is sufficient to keep V_2 cut off. When the modulator V_1 is cut off by a negative trigger pulse from the trigger unit the bias on V_2 from R_{L1} is lifted for the duration of the pulse and V_2 conducts heavily for this period. High power oscillations at the frequency of the tuned circuit are thus developed for the duration of the trigger pulse and these are applied as output to the aerial.

Fig 10 illustrates how *anode modulation* may be achieved by using a hard valve modulator. The modulator valve is placed *in series* with the high power r.f. oscillator *across* the h.t. supply. The modulator is normally held cut off by a fixed bias. The oscillator circuit has therefore no earth return and no oscillations occur until a trigger pulse from the master timing unit lifts the bias on the modulator valve. The modulator valve then cuts on for the duration of the trigger pulse and completes the circuit to the oscillator so that oscillations occur for the required period of time.

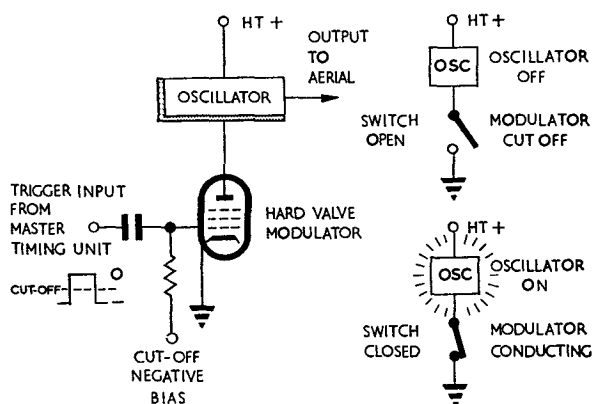


FIG 10. BASIS OF ANODE MODULATION IN A PULSED RADAR TRANSMITTER

Unlike the grid-modulated transmitter of Fig 9, the anode modulator must be capable of handling the high peak current required by the oscillator. The modulator of Fig 10 must therefore be a high power valve. The main limitation of the hard valve modulator is the difficulty of producing valves capable of handling the large power required. For this reason gas-filled valves are often used in the modulator stage of high-power radar transmitters.

Thyratron Switch

The thyratron is a gas-filled triode or tetrode (hydrogen gas is often used) which is used in many radar modulators to provide the high-power, fast-acting switch to turn the transmitter valve on and off. The main properties of a thyratron are:

a. Once the valve has started to conduct, the gas in the valve ionizes; the grid then has no further control over the anode current and the only way of switching the valve off is to reduce the anode voltage *below* a given level. Thus although a trigger voltage to the grid of a

thyatron can *initiate* a pulse it cannot end it. This is a disadvantage compared with a hard valve modulator.

b. When the valve is conducting it has a *very low* internal resistance.

c. It can pass currents of the order of amperes.

Fig 11a shows the basic block schematic outline of one type of modulator used in radar transmitters. The energy for the pulse, which is to be dissipated eventually in the load (e.g. a magnetron oscillator), is obtained from the energy source—a high-power d.c. supply. This energy is stored in the storage device at a *slow rate* during the periods between pulses, the charging impedance deciding the *rate* at which energy is supplied to the storage device. At the required instant of time the switch is closed and the stored energy is *rapidly dissipated* in the load. Since a high level of energy is being expended in a very short period of time, a pulse of very high peak power is produced ($\text{Power} = \frac{\text{Energy}}{\text{Time}}$).

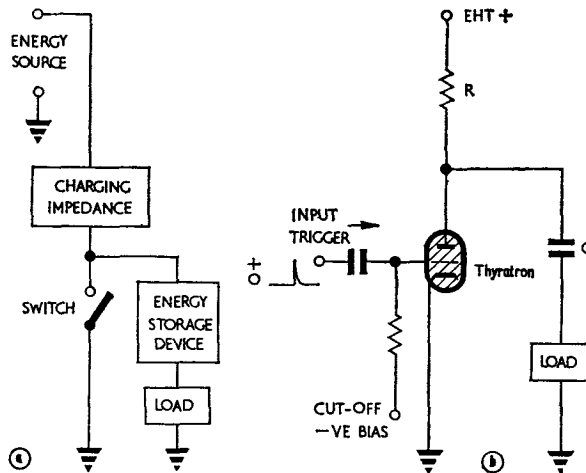


FIG 11. BASIC THYRATRON MODULATOR

A basic arrangement using this system is outlined in Fig 11b. The thyatron is cut off during the periods between pulses and during this time C charges at a relatively slow rate through R and the load. When the trigger input pulse arrives the thyatron 'fires' and, because of its low resistance, the capacitor discharges rapidly through the load and the thyatron. The energy acquired by C during the period between pulses is rapidly dissipated in the load and a high-power, short-duration pulse is produced. The capacitor discharges until the anode voltage of the thyatron is such that the valve cuts off (about +15V). When this happens, C again commences to charge through R towards the e.h.t. supply and after a short time the circuit is again ready for the next input trigger pulse.

A capacitor has the disadvantage of discharging exponentially with time and, because the thyatron cannot be cut off at its grid by the trigger pulse, the pulse developed across the load has a poor shape. For this reason a capacitor is not normally used as the energy storage device. In most radar equipments a *delay line* or *pulse-forming network* is used instead. The delay line is considered in detail in Section 5. All we need to know at present is that it can be used in much the same way as the capacitor of Fig 11 but, compared with the capacitor, it produces a much better shaped pulse of higher energy content. The pulse produced by a delay line modulator has steep leading and trailing edges and an accurate pre-determined pulse duration.

Ignitron

Other gas-filled valves beside the thyatron are used as switches in high-power radar modulators. These include the *ignitron* and the *trigatron*. In very high-power radars the modulator arrangements considered so far would require very high applied direct voltages. The same high peak power could be obtained by using a d.c. supply of a few hundred volts but supplying several hundred amperes of current. Where currents of this magnitude have to be switched on and off it is usual to use an ignitron.

The ignitron is a mercury pool switch valve. Its basic arrangement is sketched in Fig 12. It consists of a glass envelope in which there is a carbon anode and a control grid and a pool of mercury which acts as the cathode. Before the valve can operate, an arc must be 'struck' and maintained. This requires two additional electrodes, a striker electrode and an exciter anode. A voltage exists between the striker electrode and the pool of mercury and when the striker electromagnet is energized it moves the striker into the pool. This action short-circuits the electromagnet supply so that the striker arm is released and moves back under the tension of a spring to its original position. As it is lifted out of the pool an arc is created and free ions are made available for the main switching function. The exciter anode then takes over to *maintain* this arc and the striker electrode supply is removed.

Because we now have a plentiful supply of free ions already available within the envelope, when a positive trigger pulse is applied to the control grid the valve 'fires' immediately and we have a large flow of current between the main anode and the pool. Thus the ignitron may be used in much the same way as a thyatron to provide a rapid discharge of the delay line through the load. Compared with the thyatron the ignitron is capable of switching very much higher currents.

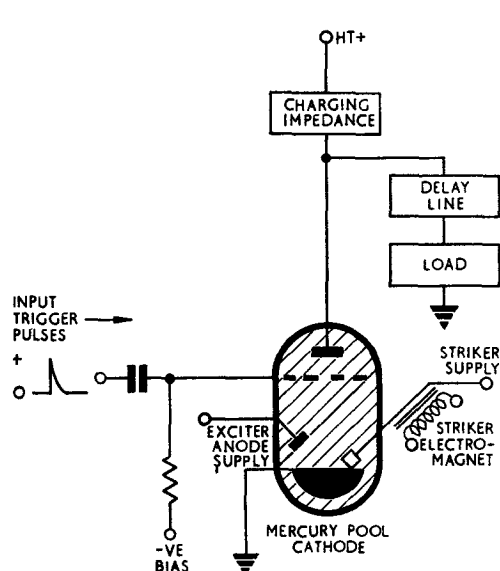


FIG 12. BASIC IGNITRON

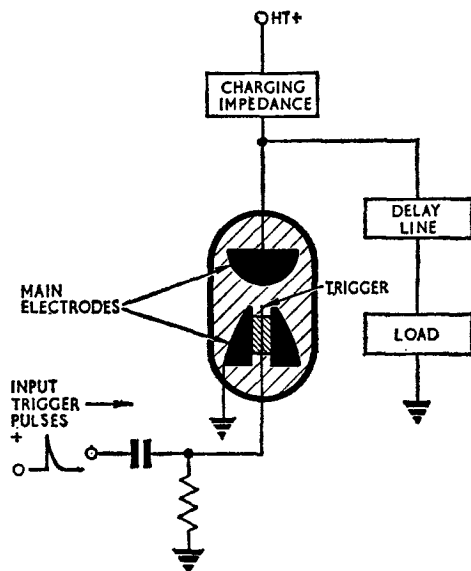


FIG 13. TRIGATRON

Trigatron

The trigatron is a form of cold-cathode gas-filled triode. Its basic construction is outlined in Fig 13. The two main electrodes are hemispherical caps, one of which has a hole in the middle into which is inserted a third electrode known as the *trigger* electrode. The spacing

between the two main electrodes is such that ionization does not occur even when the maximum available voltage is applied across them. Ionization is initiated by the application of a trigger pulse to the trigger electrode. The gap between the trigger electrode and the electrode into which it is inserted then breaks down and ionization occurs. Once the gas has been ionized in this way the voltage between the main electrodes is sufficient to spark the main gap and the valve 'fires', its resistance falling to a very low value. The spark is extinguished and the valve stops conducting when the voltage between the two main electrodes falls below the ionization voltage of the gas. The trigatron is used in a modulator circuit in much the same way as a thyatron, to discharge the delay line through the load. The trigatron passes currents of the order of a few amperes at high voltage levels. This is less than the current which a thyatron can handle (and *very much smaller* than the ignitron current) but the trigatron has the advantage that no heater supply is required.

FREQUENCY-DIVIDING & COUNTING CIRCUITS

Introduction

Frequency-dividing and counting circuits are both used in radar equipments. In frequency-dividing circuits (sometimes referred to as 'count-down' circuits) one output pulse is obtained for every n input pulses, where n may be any whole number between one and about ten. The frequency divider thus produces an output which is an exact *sub-multiple* of the pulse train which causes the circuit action. As we shall see, this has many applications in radar. A counting circuit is normally used to provide an output which is proportional to the *number* of input pulses applied to it. This too has many applications in radar. We shall consider both types of circuit in this chapter.

Frequency Division

We know from previous work that both the astable multivibrator and the free-running blocking oscillator may be *synchronized* by applying appropriate sync pulses to the control grid (valve version) or to the base-emitter circuit (transistor version). The p.r.f. of the sync pulses must be slightly *higher* than that of the free-running circuit (Fig 1) and their amplitude must be such that they lift the circuit into conduction *before* it would *normally* cut on. The output is then 'locked' to the sync pulses.

Frequency division is merely an extension of this idea. The p.r.f. of the input pulses is now made *much higher* than that of the free-running circuit and the circuit constants are then adjusted such that, say, every *third* input pulse brings the circuit prematurely into conduction. In this case one output pulse is produced for every three input pulses and the circuit is said to have a *count-down ratio* of 3 : 1. Count-down ratios of up to about 10 : 1 may be obtained in this way.

Application of Frequency Division

Frequency dividers have many applications. We shall consider only one at this stage. We saw in an earlier chapter that cal pips may be used to calibrate a type A display. A display with large cal pips at ten-mile intervals and small cal pips at five-mile intervals may be required. A cal pip generator, consisting of a triggered blocking oscillator, may be used to provide the five-mile cal pips (Fig 2). The output from this is applied to the display and also as the input to a frequency divider. This may be a free-running blocking oscillator so adjusted that every *second* input pulse from the cal pip generator brings the divider into conduction slightly earlier

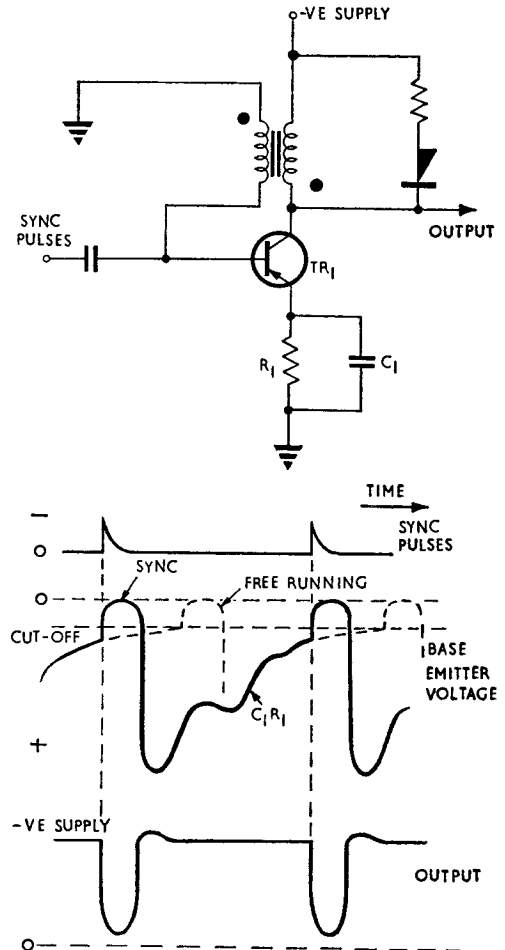


FIG 1. SYNCHRONIZED BLOCKING OSCILLATOR

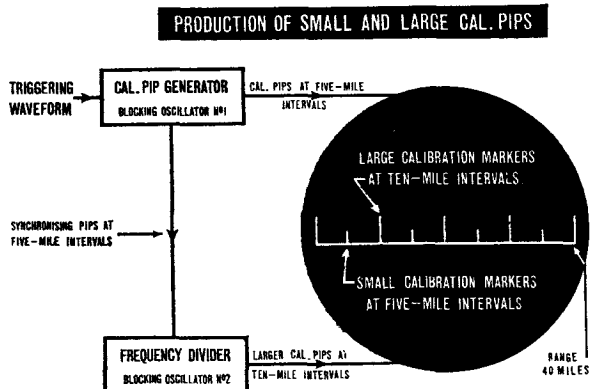


FIG 2. USE OF FREQUENCY DIVIDER IN CALIBRATED DISPLAY

than normal. The frequency divider therefore produces *one* output pulse for every *two* input pulses so that its frequency is *half* that of the input. This circuit therefore produces ten-mile cal pips.

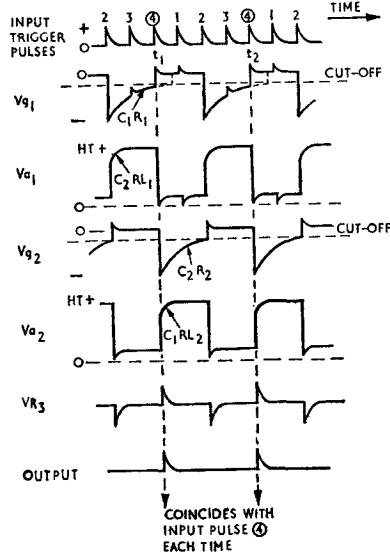
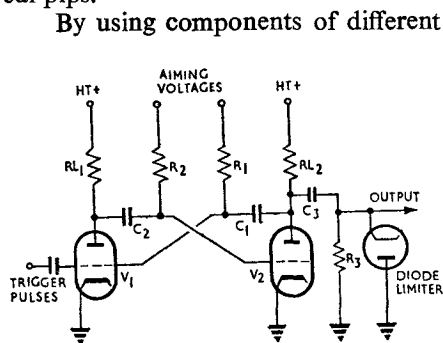


FIG 3. ASTABLE MULTIVIBRATOR AS FREQUENCY DIVIDER

By using components of different values in the two blocking oscillators we can make the ten-mile cal pips from the frequency divider larger than the five-mile cal pips from the cal pip generator. When they are both applied to the c.r.t. the display is as shown in Fig 2. By using frequency dividers in this way we can obtain an elaborate calibration display where the trace itself is marked off like a ruler in whatever units of range we require. For example, a unit which supplies 5-, 10- and 50-mile cal pips may use three separate circuits, each circuit being triggered by the one that precedes it. The 10-mile circuit operates at half the frequency of the 5-mile circuit and the frequency of the 50-mile circuit is one-fifth that of the 10-mile circuit.

Frequency Division in Multivibrator

The basic circuit of an astable multivibrator used as a frequency divider is illustrated in Fig 3. The positive-going input trigger pulses are applied to the grid of V_1 but these have no effect until the grid voltage has risen almost to cut-off. Pulse 4 at time t_1 causes V_{g1} to rise above cut-off as shown in Fig 3 and initiates the usual multivibrator avalanche. With V_1 conducting V_{a1} falls almost to zero volts and remains there whilst V_{g2} is rising to cut-off in the normal manner. When V_{g2} reaches cut-off V_{a2} falls towards zero; V_{g1} is then driven below cut-off and V_{a1} rises to h.t.+. Pulses 1, 2 and 3 have *no effect* on the circuit action, pulses 1 and 2 because V_1 is already conducting when they appear and pulse 3 because V_{g1} is too far below cut-off to be affected. Pulse 4 however at time t_2 again raises V_{g1} above cut-off *before it would normally do so* and the action is repeated. The

output from V_2 anode is differentiated by C_3R_3 and then negatively-limited by the diode. Fig 3 shows that the system delivers *one* output pulse coincident with every *fourth* input pulse. The circuit is thus counting down by a ratio of 4 : 1.

It will be noted that one half-cycle of each cycle of the multivibrator output is still free-running. This may give rise to instability if the circuit conditions change to cause V_{s2} to reach cut-off earlier or later than the required instant of time. If V_{s2} reaches cut-off earlier than normal then V_1 may be triggered by pulse 3. If V_2 cut-on is delayed, V_1 may not be triggered until pulse 1 in the next period. To improve the stability the trigger pulses may be applied to *both* grids and the circuit adjusted such that pulse 2 triggers V_2 and pulse 4, V_1 . In this way we ensure that the triggering times remain stable. For an *even* count-down ratio it also ensures that the output square wave is *symmetrical*. This may be a requirement.

This circuit is capable of dividing by up to as much as ten times. At higher count-down ratios the exponential rise of the grid voltage becomes too 'flat' near cut-off to ensure reliable switching by the correct trigger pulse (Fig 4). Increasing the positive aiming voltage may improve this but the available time period is then reduced so that this in itself limits high count-down ratios.

We know that an astable multivibrator will continue to free-run in the absence of trigger pulses. This is a disadvantage in some circuits because it is often better that failure should cause a null output than an incorrect one. When this is the requirement a *monostable* multivibrator (flip-flop) may be used as the frequency divider (Fig 5). The flip-flop reverts to its stable state after a relaxation period and will not operate in the absence of trigger pulses. In Fig 5 TR_2 is normally conducting and TR_1 is cut off by the positive bias applied to its base. The input applied to the trigger circuit (C_1 , R_1 , D_1) causes negative-going trigger pulses to be applied to TR_1 base. The first pulse (pulse number 5) cuts on TR_1 and the usual avalanche occurs to switch off TR_2 . V_{C1} therefore falls to just above zero volts and V_{C2} rises to just below the negative

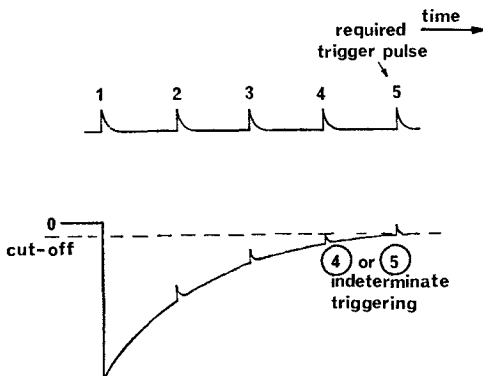


FIG 4. EFFECT OF EXCESSIVE COUNT-DOWN RATIO

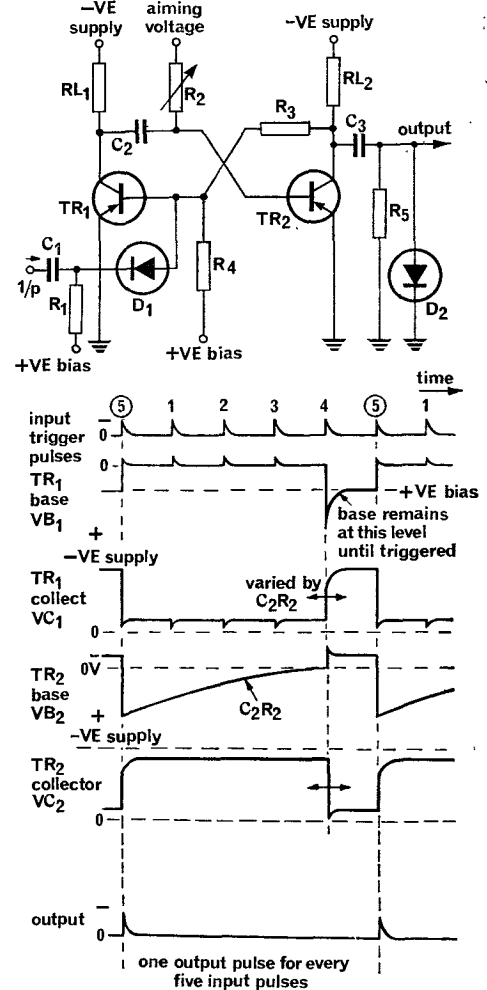


FIG 5. FLIP-FLOP AS A FREQUENCY DIVIDER

supply voltage level (determined by the values of R_{L2} , R_3 , R_4). The circuit stays in this condition whilst C_2 is discharging through R_2 causing V_{B2} to rise towards cut-on. Any trigger pulses appearing during this time have *no effect* since TR_1 is already conducting. When V_{B2} reaches cut-on the circuit quickly reverts to its stable state in which TR_1 is held cut-off by the bias voltage and TR_2 is conducting. It *remains* in this condition until the next trigger pulse appears, when the action is repeated. The count-down ratio in Fig 5 is 5 : 1.

Frequency Division by Blocking Oscillator

The circuit and waveforms of a blocking oscillator used as a frequency divider are shown in Fig 6. The circuit illustrated uses collector-base coupling and base timing (see p 152). The free-running frequency of the blocking oscillator, set by the aiming voltage on C_B , is slightly *lower* than the required *output* frequency. The p.r.f. of the trigger pulses is made *much higher* than that of the free-running circuit and the conditions are then such that, in the example shown, every *fifth* trigger pulse raises the base voltage above cut-off earlier than normal. This gives a count-down ratio of 5 : 1.

Frequency Division by Thyatron

The circuit and waveforms of a thyatron frequency divider are shown in Fig 7. To cause the valve to strike, a trigger pulse must be applied to the grid. On receipt of trigger pulse 3 at time t_1 , the valve cuts on, the gas ionizes and a high value of anode current flows. This causes V_a to fall sharply. The current flowing through the valve charges C_K causing the cathode to rise exponentially. A point is reached where the current through the valve is insufficient to maintain ionization and the valve cuts off. The anode then returns to HT+ and the cathode starts to fall as C_K discharges through R_K . Pulses 1 and 2 have no effect on the circuit action but by the time pulse 3 arrives at t_2 , V_K has fallen sufficiently to allow the valve to ionize again and the action is repeated. In the example given the count down ratio is 3 : 1 but this may be varied either by changing the time constant $C_K R_K$ or by changing the amplitude of the trigger pulses applied to the grid by varying R_1 .

Frequency Division by Tunnel Diode

The basic circuit and waveforms of a tunnel diode used as a frequency divider are shown in Fig 8. The tunnel diode is biased to point x on the negative resistance portion of its characteristic (Fig 8b). Any slight variation of current produces a back e.m.f. across the inductance L and this back e.m.f., superimposed on the bias voltage, varies the operating point on the characteristic and causes the circuit to oscillate. Let us see how this happens.

We shall assume that the circuit is operating at point A on the characteristic and that the voltage is increasing towards the bias level. The current will tend to *decrease* as the negative resistance region is entered. The inductance will, however, *maintain* the current through the diode *nearly constant* and a back e.m.f. will be induced across L in the process. This back e.m.f. is superimposed on the bias voltage and, since the current is practically constant, the operating point is transferred almost instantaneously to B on the characteristic. The back e.m.f. now falls at a rate determined by the time constant of the circuit and so does the current, until point C is reached. At C the current tends to *rise* again and once more a back e.m.f. is induced in L which holds the diode current nearly constant. Since the current is tending to rise, the back e.m.f. now *opposes* the bias voltage and, with current constant, the operating point is transferred to D on the characteristic. The current in the tunnel diode now rises normally through E to point A and the action is repeated.

When used as a frequency divider, trigger pulses are applied as shown in Fig 8c. Any trigger pulses applied during the time intervals D to E and B to C are effectively suppressed because the slope resistance of the tunnel diode is very small compared with the resistance R in these

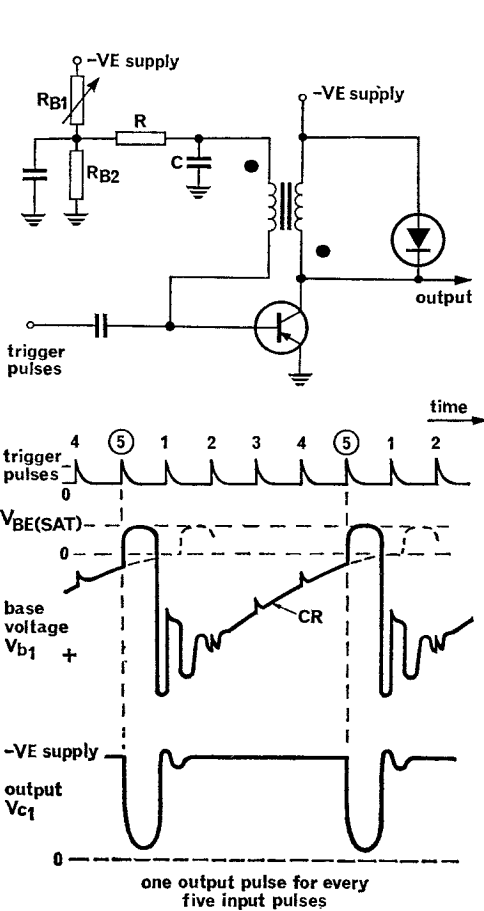


FIG 6. BLOCKING OSCILLATOR AS A FREQUENCY DIVIDER

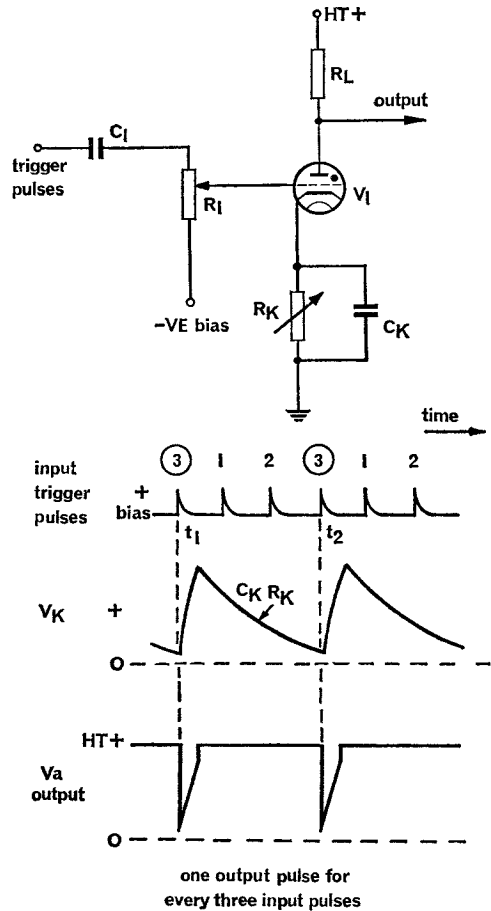


FIG 7. THYRATRON AS A FREQUENCY DIVIDER

regions. Between E and A however the diode resistance *increases* sharply and a trigger pulse applied during this time 'locks' the circuit before point A is reached. The circuit shown is triggered by every *fifth* input pulse, giving a count-down ratio of 5 : 1. *Very high* count-down ratios of up to 40 : 1 may be achieved with this circuit.

Counting Circuits

Counting circuits have two main uses in radar:

- As a means of measuring the pulse repetition frequency of a circuit, *ie* the number of regular pulses occurring *per second*.
- As a means of counting the number of *random* pulses which have occurred over a *given* period of time.

A number of circuits have been devised for these applications, including bistable multi-vibrators and gas discharge valves, but the most usual circuit arrangement uses diodes.

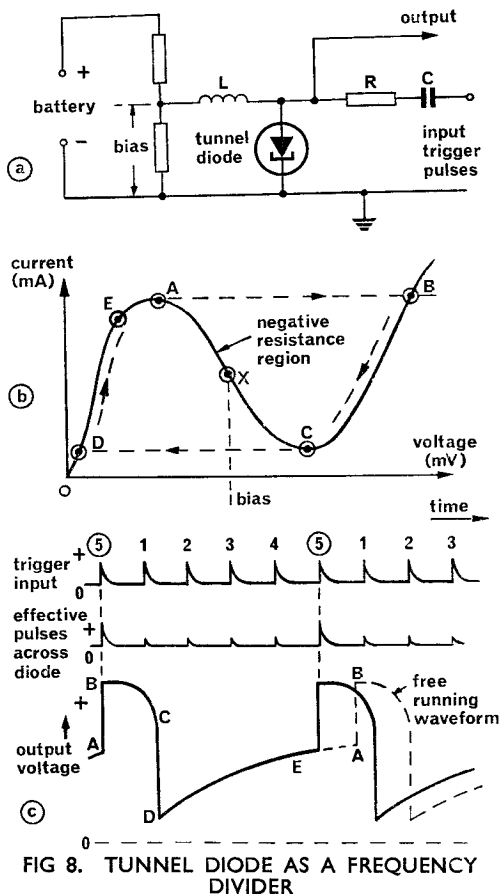


FIG 8. TUNNEL DIODE AS A FREQUENCY DIVIDER

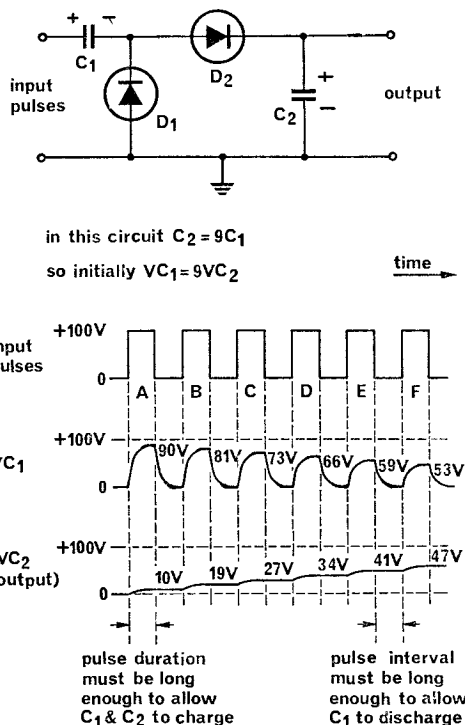


FIG 9. DIODE STEP-BY-STEP COUNTER

Diode Step-by-step Counters

The circuit and waveforms of a basic diode step-by-step counter (sometimes referred to as a diode 'pump' circuit) are illustrated in Fig 9. A train of positive-going input pulses is applied to the circuit. The leading edge of the first input pulse at A is applied through C_1 to D_2 causing the diode to conduct. Both C_1 and C_2 then charge rapidly through D_2 until $V_{C1} + V_{C2} = \text{Applied Voltage (100V)}$. The capacitor C_1 is usually small compared with C_2 . If $C_2 = 9C_1$ (typical figures) then one-tenth of the applied voltage appears across C_2 (in this example 10V) and nine-tenths is developed across C_1 (in this example 90V).

At the end of the pulse the applied voltage falls to zero and the voltages across C_1 and C_2 are then such as to cut D_2 off. C_2 therefore *remains charged* to +10V. On the other hand the voltage across C_1 is of the correct polarity to cut on D_1 and the diode rapidly *discharges* C_1 to zero. If D_1 were not present C_1 would also remain charged and there would then be no further circuit action.

The leading edge of the next input pulse at B again cuts on D_2 , but since C_2 is already charged to +10V the *effective* input voltage is now the *difference* between the applied voltage and V_{C2} , *ie* 90V. This new voltage is again shared between C_1 and C_2 in the ratio one-tenth across C_2 and nine-tenths across C_1 (assuming $C_2 = 9C_1$). Thus C_1 charges to 81V and C_2 rises by 9V to +19V.

During the next pulse interval B-C, D_2 again cuts off to hold V_{C_2} at +19V, and D_1 conducts to discharge C_1 to zero.

Thereafter, the action is repeated. Each input pulse increases V_{C_2} in steps but the increase in amplitude of successive steps is becoming progressively less. The waveform of voltage across C_2 is often referred to as a 'staircase' waveform and has been used in some circuits. The staircase does not rise linearly however; because of the progressively smaller steps the rise is exponential.

Note that after a large number of cycles, V_{C_2} would eventually reach 100V and the circuit action would then cease. In practice however C_2 is discharged by another circuit before the voltage steps become so small that it is difficult to distinguish between them. When C_2 is discharged the circuit returns to its initial state and the action may be repeated indefinitely.

One typical circuit for the discharge of C_2 after a given number of input pulses is shown in Fig 10. The circuit is that of a triggered blocking oscillator which is normally biased beyond cut-off by the potential divider R_1R_2 between h.t.+ and earth. The bias may be adjusted by R_2 and it is set such that after a given number of input pulses, V_{C_2} has risen sufficiently to lift V_3 grid above cut-off. Normal blocking oscillator action then takes place, the grid current of V_3 discharging C_2 to zero ready for the next series of input pulses. We thus obtain one output pulse

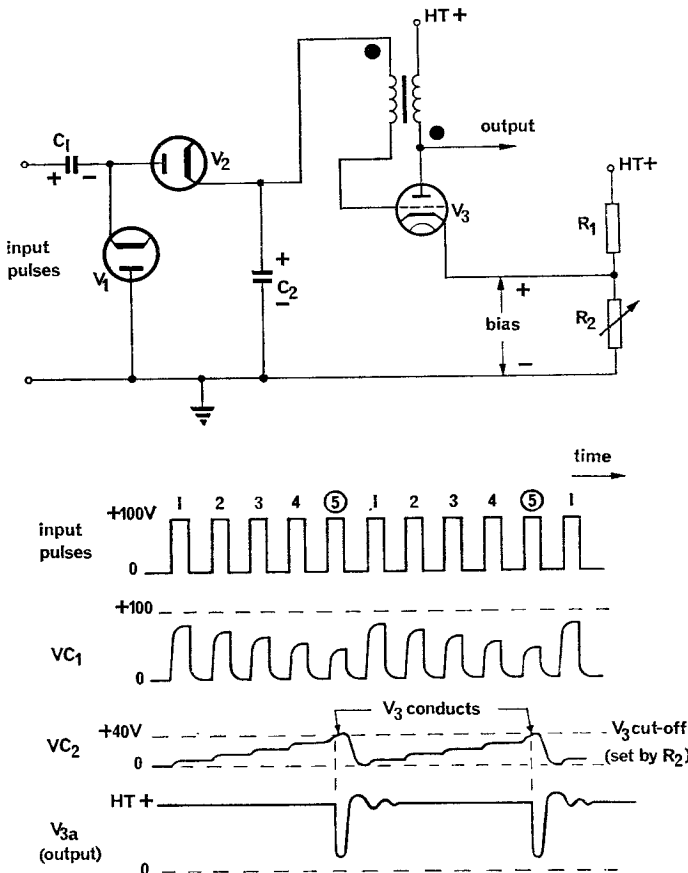


FIG 10. APPLICATION OF STEP-BY-STEP COUNTER

after a *given number* of input pulses. In the example shown every *fifth* input pulse triggers the blocking oscillator. Each output pulse thus indicates a count of five input pulses. In this particular application the counter is doing a similar job to the dividers considered earlier in this chapter. The important difference is that the input to the counter may be *random* pulses occurring at irregular instants of time, because C_2 cannot discharge between pulses.

Although the point at which triggering of V_3 occurs may be varied by R_2 (and also by the ratio of C_1 to C_2) the circuit is not normally expected to count more than about ten pulses before C_2 is discharged. The reason for this is clear when it is remembered that the steps at the top of the 'staircase' are very shallow. It becomes more difficult to distinguish between adjacent steps so that the triggering becomes less determinate.

There are a number of variations of this circuit. The 'discharge' circuit need not be a blocking oscillator; thyatron and Miller integrator circuits have also been used. A *negative-going* output across C_2 may be obtained by reversing the polarity of the two diodes V_1 and V_2 .

Transistor Counter

The main limitation of the diode step-by-step counter is the non-linear rise of the staircase waveform. This may be overcome by the circuit of Fig 11 in which the diode D_1 is replaced by the n-p-n transistor TR_1 , the output voltage at Y being fed back direct to the transistor base. The operation of the circuit is basically similar to that of the diode counter except that TR_1 conducts during the periods between pulses because of the positive voltage at its base from C_2 . With TR_1 conducting, C_1 discharges until V_{C_1} equals $-V_{C_2}$. (See p 200 on 'bootstrapping'). The result is that D_2 has practically no reverse bias to overcome so that the whole input voltage is available during each pulse to charge C_1 and C_2 —unlike the diode counter where the effective input is the difference between the applied voltage and V_{C_2} . Thus, in Fig 11, the voltage steps of V_{C_2} are all equal and the staircase waveform becomes linear.

PRF Meter

The circuit of a transistor counter used as a p.r.f. meter is shown in Fig 12a. This circuit is similar to the step-by-step counter except that C_2 is now shunted by R_1 . The result is that during each interval between pulses C_2 can discharge via R_1 . V_{C_2} will rise in the manner described for the step-by-step counter and R_1 will discharge C_2 in the intervals. This continues until the charge *gained* via D_2 during the pulse is exactly equal to that *lost* via R_1 in the interval between pulses (Fig 12b). If the p.r.f. *increases*, the charge gained by C_2 increases; V_{C_2} therefore rises until a new (higher) balance is obtained. The *mean* voltage across C_2 is therefore proportional to the p.r.f. The ripple component of the output is smoothed by R_2C_3 and the resulting d.c. voltage is applied as the input to the base of the n-p-n transistor TR_2 . As the p.r.f. of the input increases, V_{C_2} rises causing TR_2 base voltage to rise. This increases the current through TR_2 and causes a higher reading in the meter which is calibrated to give a direct reading of the p.r.f. R_3 varies the bias to TR_2 base for calibration purposes. To obtain an output *voltage* proportional to the p.r.f. of the input, a collector load may be inserted in TR_2 and the output voltage taken from the collector.

Bistable Counters

Various other circuits may be used as counting circuits. Apart from those discussed in this chapter the one most frequently used, especially in computers, is the bistable multivibrator. The Eccles-Jordan bistable is a natural 'scale-of-two' (binary) counter because to go through *one* cycle of operation *two* trigger input pulses are required. Thus each output pulse indicates two input pulses.

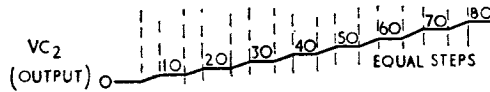
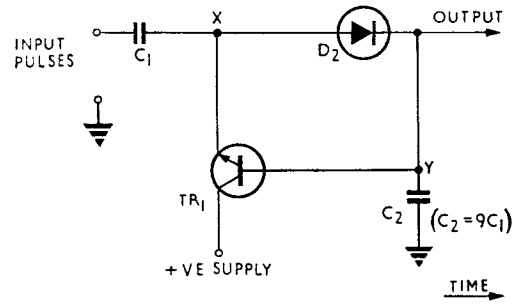
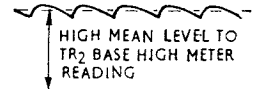
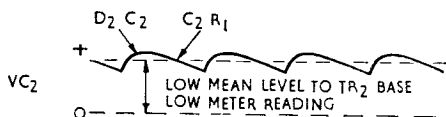
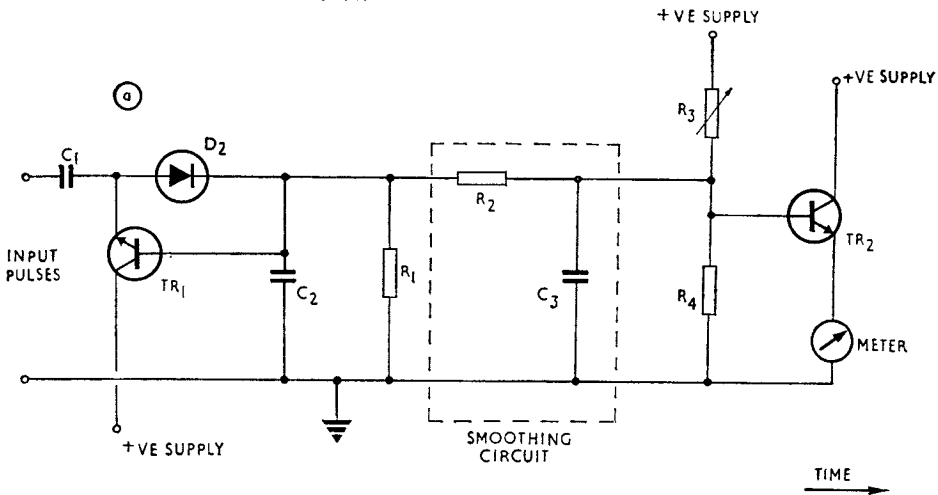


FIG 11. TRANSISTOR COUNTER



(b)

LOW PRF

HIGH PRF

FIG 12. CIRCUIT FOR INDICATING PRF

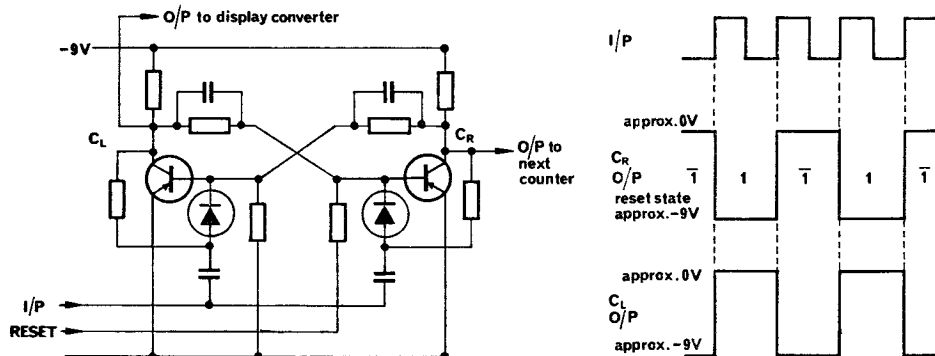


FIG 13. ECCLES-JORDAN BISTABLE AS A BINARY COUNTER

Bistable as a binary counter. The basic circuit action of an Eccles-Jordan bistable was discussed on p 127. Inclusion of two diodes in the base input leads (Fig 13) means that the circuit will only change states on a positive going edge of the trigger input, *ie* once per cycle. Thus for two cycles of input the bistable will change states twice, producing one cycle of output. This means that the circuit is dividing or counting by two and it is therefore called a binary counter. Such bistable circuits may be used in combination in a 'register' to provide counts in *powers* of two. With two bistables connected in cascade, one output cycle indicates a count of $2^2 = 4$; for three circuits in cascade an output is obtained for every $2^3 = 8$ input cycles and so on.

Before the count begins all bistables are reset by momentarily open-circuiting the reset line (normally earthed). This causes a negative voltage at the base of the right-hand transistor which therefore conducts and its collector will then be at approximately zero volts. A positive edge is produced at the output every time the bistable changes over from 1 to $\bar{1}$ and this is used to trigger the next bistable in a cascade counter.

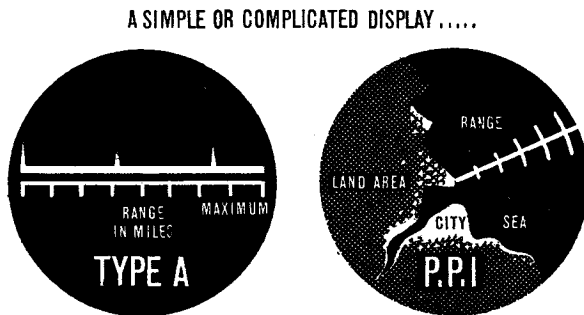
By introducing suitable feedback, a four stage cascaded counter can be made to count by ten instead of sixteen. This is a very common arrangement in computers and in test equipment such as frequency counters.

CHAPTER 12

TIMEBASE PRINCIPLES

Introduction

In Section 1 we saw some of the ways in which received information is displayed on a radar indicator. The type of display we use is determined by the job which the equipment is designed to do, but no matter how complicated the display may seem it is always built up by *one spot* of light moving rapidly over the screen of the c.r.t. (Fig 1).



..... IS BUILT UP BY ONE SPOT MOVING ACROSS THE SCREEN

FIG 1. TYPICAL RADAR DISPLAYS

In this chapter we are going to learn *how* the spot is made to move in the required manner. Before we do so however we shall recall a few essential facts about the action of a c.r.t. (see AP 3302 Part 1B p 515).

Any c.r.t. can be divided into four main parts. These, together with their function, are illustrated in Fig 2. If the c.r.t. uses electric fields for deflecting and focusing the electron beam, we have an *electrostatic* c.r.t. A *magnetic* c.r.t. uses external coils to provide focusing and deflection. Some c.r.t.s use a combination of electrostatic focusing and magnetic deflection.

Controls are also provided so that the appearance of the picture on the screen may be adjusted. These include:

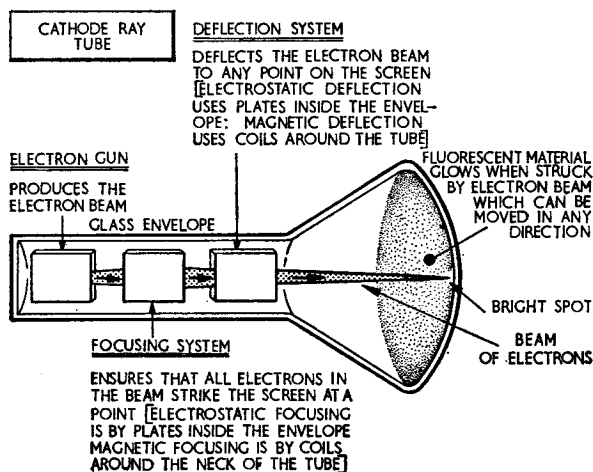


FIG 2. MAIN PARTS OF A CRT

a. *Brilliance.* This control enables the operator to adjust the *brightness* of the display to the most suitable level. Varying the control alters the bias between the control grid and the cathode of the c.r.t. thus altering the intensity of the electron beam.

b. *Focus.* This control is set to focus the spot to the smallest possible size, otherwise the display appears blurred and the target indications are not distinct.

Type A Display

We met the type A display in Section 1. This was the first display to be used in radar equipment and is probably the simplest to produce. We shall therefore take this as the first example and see what we require the spot to do. We can then go on to find out how the desired results are obtained.

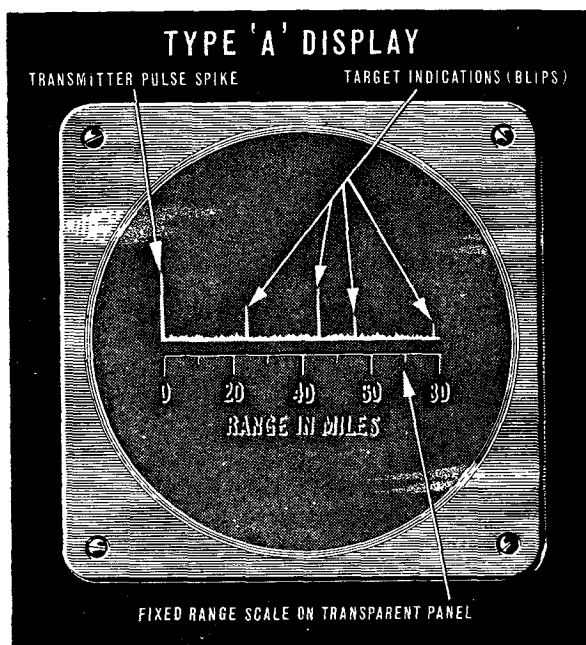


FIG 3. TYPE 'A' DISPLAY

As shown in Fig 3 the spot moves across the screen to form a *horizontal trace*, and the reflected pulses from distant targets deflect the spot upwards to give blips on the trace. A range scale is fitted against the face of the c.r.t. so that the operator can read the range of any target by noting the position of the selected blip on the scale.

How the Spot Moves in a Type A Display

In this chapter we are concerned only with the *trace*. We shall therefore break up the display by removing the range scale and disconnecting the receiver from the indicator so that we are left with only a horizontal trace on the screen (Fig 4).

When each transmitter pulse commences, the spot must be on the left-hand side of the screen at point P. If the maximum radar range is 80 miles the spot must move along the trace at a *constant speed* to reach point Q exactly $860 \mu\text{s}$ after it leaves point P. This movement of the spot from P to Q is called the *sweep*; the time taken to travel from P to Q is determined by the maximum range of the equipment and is known as the *sweep time*.

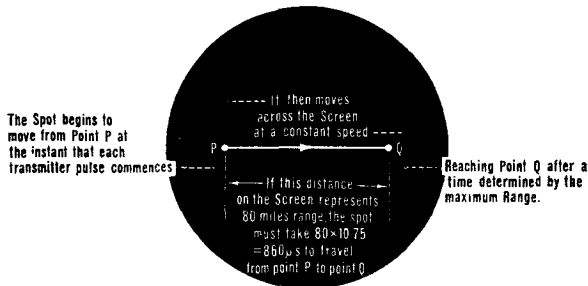


FIG. 4. SWEEP OF A TYPE 'A' DISPLAY



FIG. 5. FLYBACK OF A TYPE 'A' DISPLAY

If the distance on the screen between points P and Q is four inches, then on switching from the 0-80 miles range to, say, the 0-20 miles range, the spot must still move four inches across the screen in the course of each sweep; but now the sweep time must be $20 \times 10.75 = 215 \mu\text{s}$. Thus when we switch from one range to another we must make the spot move *at a different speed*; as the spot movement is in a specified direction we say that we change the *sweep velocity*.

At the end of each sweep the spot is on the right-hand side of the screen at point Q (Fig. 5). In the interval between the end of one sweep and the beginning of the next we must make the spot return to its starting point at P. This return movement is called the *fly-back*. We do not want to see the flyback on the screen and so must apply either a *negative-going* blanking pulse to the c.r.t. *grid* or a *positive-going* pulse to the c.r.t. *cathode* during the flyback time to cut off the electron beam and *blank out* the screen.

As we saw in AP 3302 Part 1B (p172) the movements of the spot across the screen to form the trace, and also the flyback, are repeated continuously and the repetition rate is made sufficiently high so that the eye sees the movement as a continuous trace on the screen.

Each sweep commences at the instant the transmitter fires, and to find the range of a target we really measure the *time interval* between the beginning of each sweep and the reception of the reflected pulse. By making the spot move over a definite path at a known velocity we say that we are providing a *timebase* against which the range of a target can be measured. The spot is made to move over the required path at the correct velocity by a specially-shaped waveform of voltage or current which we apply to the deflecting system of the c.r.t. The circuit which produces this waveform is called a *timebase generator*.

The actual *shape* of the timebase waveform depends upon whether an electrostatic or a magnetic c.r.t. is being used. For the moment we shall consider only the electrostatic type.

Timebase Deflection in Electrostatic CRT

The deflecting system in an electrostatic c.r.t. is shown in Fig. 6. As we already know, the electron beam is deflected from its central path by electric fields between the plates of either pair.

For certain practical reasons we usually apply the timebase waveform to the X plates and the signal to be examined to the Y plates. Let us now see what the shape of the timebase waveform must be if we wish to produce the type A display considered earlier.

The simplest way of obtaining the required timebase is to use *single-ended* deflection. This means that one of the X plates (say X_1) is connected permanently to earth and the timebase waveform is applied to the other plate X_2 . In practice, in radar systems, *balanced* deflection is normally used. In this, the timebase voltage is split into two *equal* deflecting voltages of *opposite* polarity for application to *both* deflecting plates. However, single-ended deflection is easier to understand and, although it introduces a certain amount of distortion, it is the method we shall consider here. Balanced deflection is dealt with later in Chapter 16 (p219).

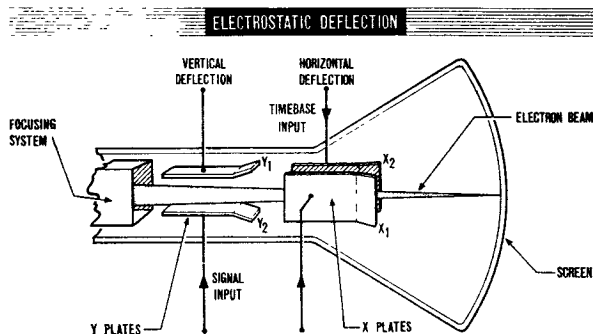


FIG. 6. DEFLECTION IN ELECTROSTATIC CRT

Let us suppose that we are using single-ended deflection in which X_1 is permanently connected to earth. When X_2 is also at earth potential there is no electric field between the plates and the spot is at the centre of the screen. To move the spot to its starting position at point P in Fig. 7 we make X_2 *negative* with respect to earth. The voltage between the plates then produces an electric field which deflects the beam to the *left-hand* side of the screen at point P. Similarly, if X_2 is *positive* with respect to earth the beam is deflected to the *right-hand* side of the screen towards point Q.

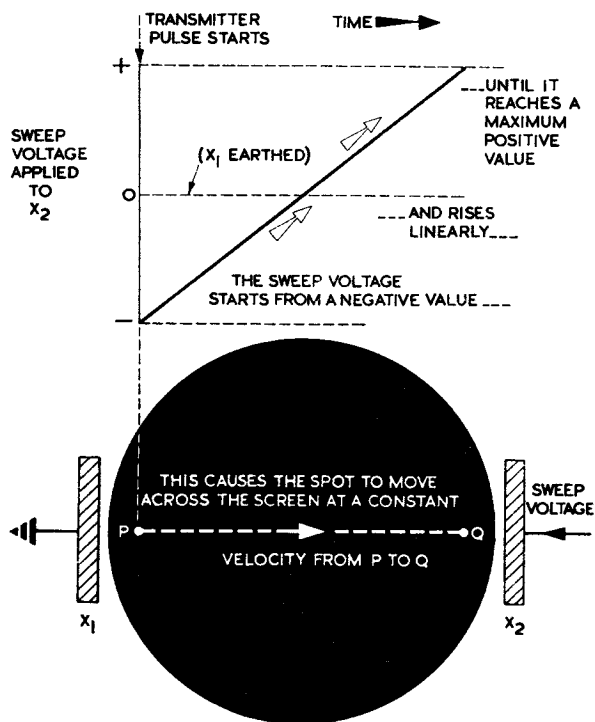


FIG. 7. PRODUCTION OF HORIZONTAL TRACE IN ELECTROSTATIC CRT

To make the spot *sweep* from P to Q, thereby producing a horizontal trace, we make the voltage on X_2 *rise evenly* from a negative to a positive value relative to earth. The effect of this is shown in Fig 7.

When the sweep has ended we must cause the voltage on X_2 to *fall rapidly* from its final positive value to its original negative value so that as quickly as possible after the sweep ends the spot is again at point P, ready for the next sweep to commence. The *shape* of the *flyback* portion of the waveform is not usually important because the c.r.t. is blanked out for the duration of the flyback time.

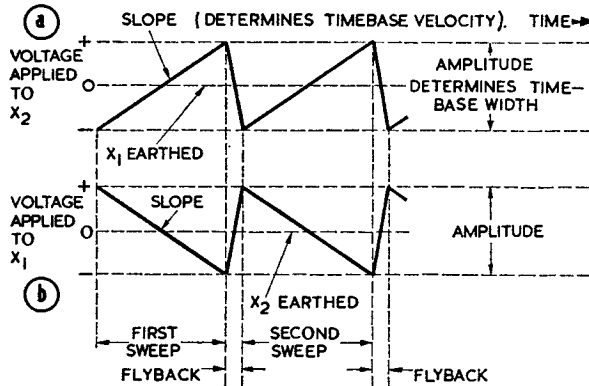


FIG 8. SAWTOOTH TIMEBASE WAVEFORMS

The complete timebase waveform which we apply to X_2 is as shown in Fig 8a. Because it resembles the teeth of a saw we call it a *sawtooth* waveform.

The same results would be obtained if we earthed X_2 and applied a *negative-going* sawtooth waveform to X_1 , as in Fig 8b. It does not matter, therefore, whether we consider positive-going or negative-going sawtooths because both produce the same result when applied to the appropriate X plate.

The *slope* of the sawtooth during the sweep time determines the *speed* with which the spot is deflected. This determines the *timebase velocity*, and the steeper the slope the greater is the velocity. The *amplitude* of the sawtooth determines the *distance* through which the spot is deflected, i.e. the *trace length* or *timebase width*.

It is usual to have a timebase which at all times occupies the full width of the c.r.t. screen so as to have as large a 'picture' as possible. This means that the sawtooth *amplitude* must always be constant.

Let us assume that the required length of trace on a radar display is four inches. We have seen earlier in this chapter that if we require to measure targets at a range of up to 20 miles

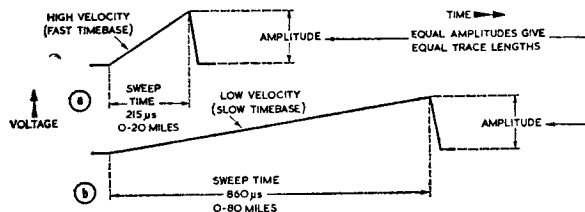


FIG 9. VARIATION OF TIMEBASE VELOCITY, AMPLITUDE CONSTANT

the duration of the sweep portion of the timebase must be $215\ \mu\text{s}$ (Fig 9a). On switching to the 0-80 miles range, the trace must still be four inches (i.e. the *amplitude* of the timebase remains the same) but the sweep time is increased to $860\ \mu\text{s}$ (Fig 9b). What we have done is to *decrease* the timebase *velocity* for the longer ranges, i.e. a *longer time* is available for the spot to cover the same length of trace. The terms 'fast' and 'slow' timebases are sometimes used to describe this.

The actual length of the trace for a given amplitude of timebase voltage depends upon the *deflection sensitivity* of the c.r.t. This is measured in millimetres or centimetres per deflection volt. If two c.r.t.s have the same timebase input the tube with the higher deflection sensitivity will give the greater deflection and the longer trace.

The timebase must also be such that the spot moves *at a constant speed* across the screen so that equal distances measured along the trace represent equal intervals of time (i.e. equal ranges in radar). To provide this, the timebase voltage must increase steadily and *linearly*. The effect of linear and non-linear timebase voltages is shown in Fig 10.

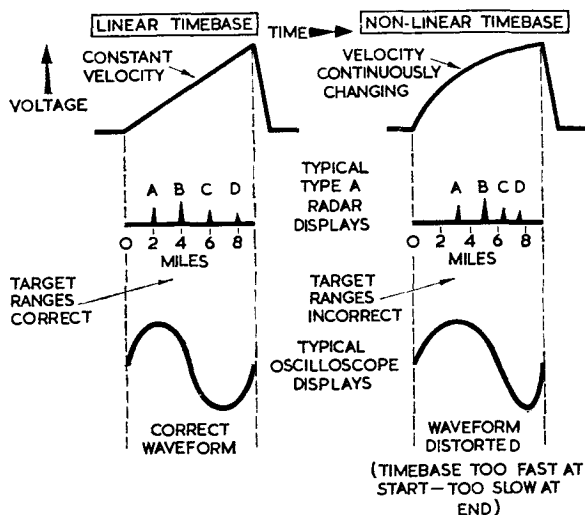


FIG 10. EFFECT OF LINEAR AND NON-LINEAR TIMEBASES

Basic Timebase Generator for Electrostatic CRT

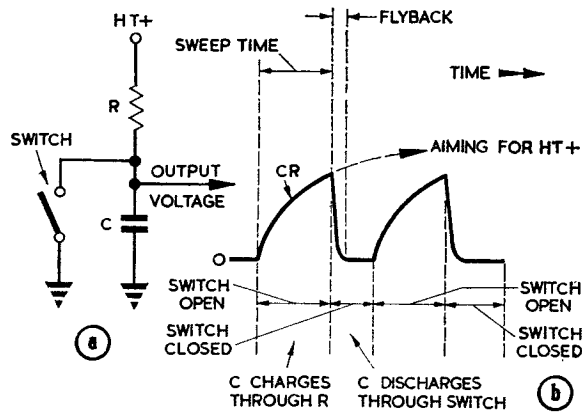
The action of most timebase generators depends upon the *slow charge* of a capacitor through a high value resistance and a *rapid discharge* through a low resistance. The basic circuit is shown in Fig 11a.

When the switch is open, the capacitor charges towards h.t.+ through the resistor, and the output voltage rises on a time constant of CR seconds. When the switch is closed, the capacitor discharges rapidly through it, and the output voltage falls to zero (Fig 11b).

Since the rise of voltage across the capacitor is exponential, a timebase produced by it would *not be linear*. However a small amount of non-linearity can usually be tolerated, and the slope of the sawtooth may be linear enough to be useful if we use only a *small part* of the capacitor voltage curve and dispense with the 'flatter' part of the curve which occurs towards the end of the charging time. This of course limits the *amplitude* of the sawtooth since the total available voltage is not then used.

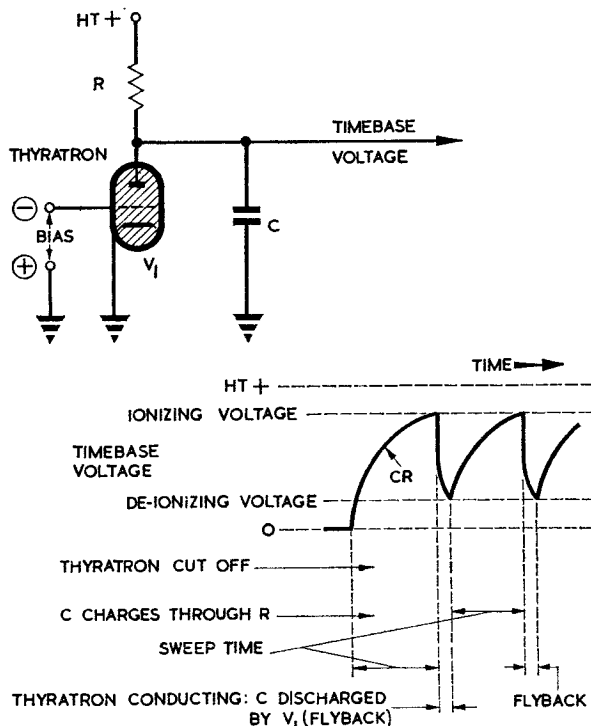
Simple Thyatron Timebase Generator

The first requirement in a practical circuit is a switch which can be operated rapidly. A *thyatron* (see p 160) is sometimes used as the switch. The main points about its operation are



that when it conducts, the gas ionizes and the valve passes a heavy current. When it does so the grid loses control and the valve can then only be cut off by the removal or the reduction of the *anode* voltage. Once the valve has cut off, the grid bias can *keep* it cut off until the anode voltage once more rises to the ionizing or striking voltage, this level being obtained by the grid bias.

A simple circuit arrangement is illustrated in Fig 12. Before the h.t. supply is connected, C is discharged and the anode voltage of the thyatron is zero. On connecting the h.t. supply,



C commences to charge through R, and the anode voltage rises; when it reaches the ionizing voltage the valve conducts. During conduction the thyatron presents a very low resistance and C discharges rapidly through it until the anode falls to the de-ionizing voltage. The valve then cuts off and is held cut off by the bias voltage. C then re-charges and the action is repeated for as long as the h.t. supply is connected. The circuit therefore acts as a *free-running* timebase generator.

Certain modifications may be made to the basic circuit to make it more practicable. These are illustrated in Fig 13a:

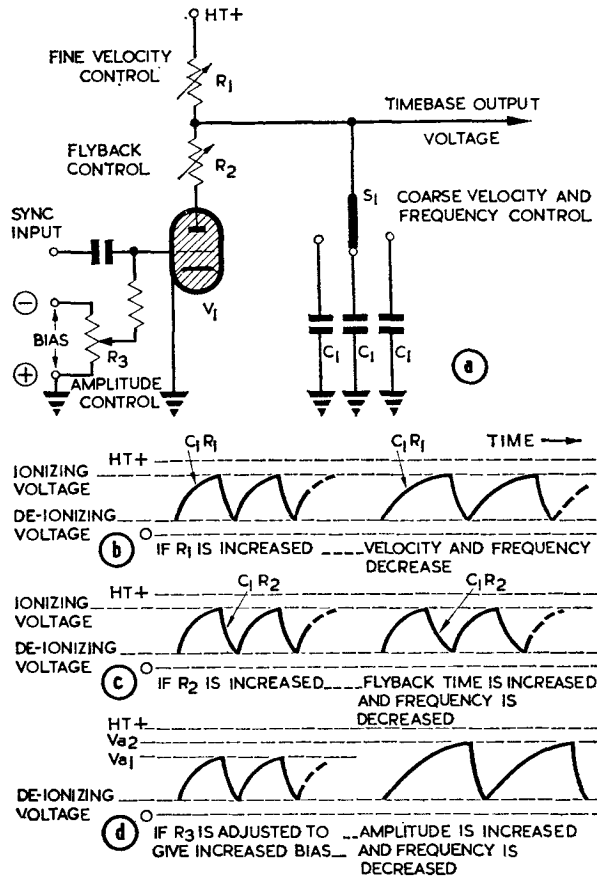


FIG 13. PRACTICAL THYATRON TIMEBASE GENERATOR AND CONTROLS

1. Varying R_1 varies the $C_1 R_1$ charging time constant and hence the *rate* at which C_1 charges. This in turn varies the *velocity* of the timebase. Since the circuit is *free-running* the *frequency* is also varied (Fig 13b). Similar results may be obtained by having several different values for C_1 to choose from; turning the switch S_1 gives a 'coarse' control of both velocity and frequency.

2. R_2 varies the time the capacitor takes to *discharge*, i.e. the flyback time, and thus provides a 'fine' *frequency* control *without affecting velocity* (Fig 13c). This control is sometimes referred to as a 'trigger' control.

3. By making the grid bias variable the anode voltage at which the valve strikes can be varied. In Fig 13d V_{a1} and V_{a2} are voltages across the thyatron and V_{a2} is larger because the grid has been made *more negative* so that a larger voltage is required to make the valve strike. This increases the *amplitude* of the sawtooth and decreases the *frequency* (since the circuit is free-running). R_3 is therefore an amplitude control.

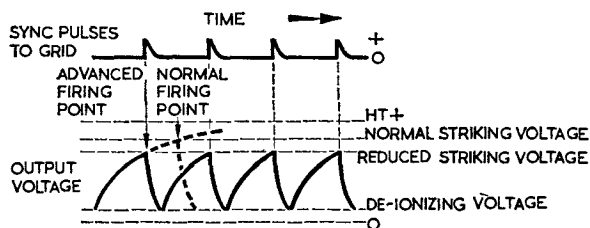


FIG 14. WAVEFORMS OF SYNCHRONIZED THYRATRON TIMEBASE GENERATOR

4. Positive *sync pulses* at a slightly higher frequency than the timebase free-running frequency may be applied to the grid. If these are applied just before the thyatron normally fires, the valve conducts earlier than normal and flyback commences. Thus the timebase frequency is 'locked' to that of the sync pulses (Fig 14).

Constant Current Charging

In the circuit described above the only attempt made to achieve linearity was to use a *small part* of the available input voltage for charging the capacitor and dispense with the 'flatter' part of the curve. A better solution would be to remove the curve in the waveform completely.

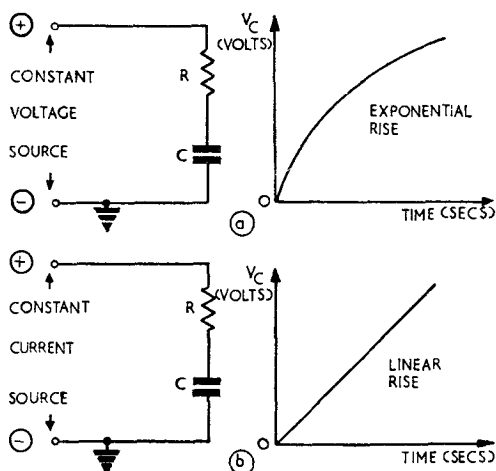


FIG 15. EXPONENTIAL AND LINEAR CHARGE OF CAPACITOR

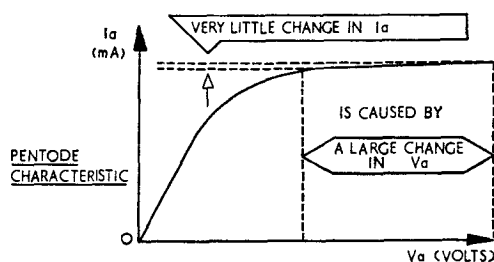


FIG 16. PENTODE AS A CONSTANT CURRENT SOURCE

The reason for the curved waveform is obvious when we remember that when C is charged through R from a source of *constant voltage*, V_C rises *exponentially* (Fig 15a). If we could make the charging *current* flow into C at a *constant rate*, V_C will also change at a constant rate and will rise *linearly* (Fig 15b).

A pentode valve is a good *constant current* source because, as shown in Fig 16, over a large part of the characteristic the anode voltage can change considerably without affecting the current

to any great extent. Let us see how we can use this effect to obtain a *linear* waveform of voltage across C .

In Fig 17, before we close the switch, C is uncharged. Thus when we switch on, the full h.t. supply of $+300\text{V}$ is applied through the uncharged capacitor (which cannot charge instantly) to the anode of V_1 .

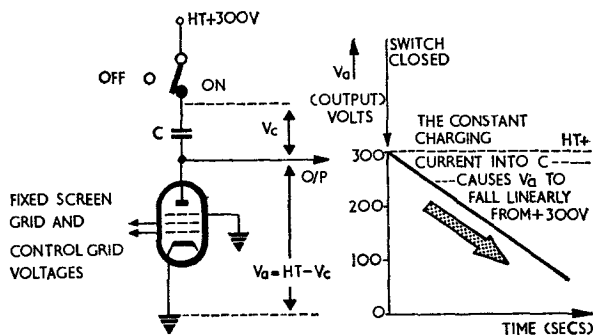


FIG 17. CONSTANT CURRENT CHARGING OF CAPACITOR

The valve then conducts and C commences to charge through the resistance of the pentode. This causes V_a to fall but, as we have seen, the fall in V_a has negligible effect on I_a in a pentode. Thus the *constant charging current* into C causes V_c to rise linearly. V_a therefore *falls linearly* from $+300\text{V}$ as shown in Fig 17. This linear fall of voltage can be used as the timebase *sweep* voltage.

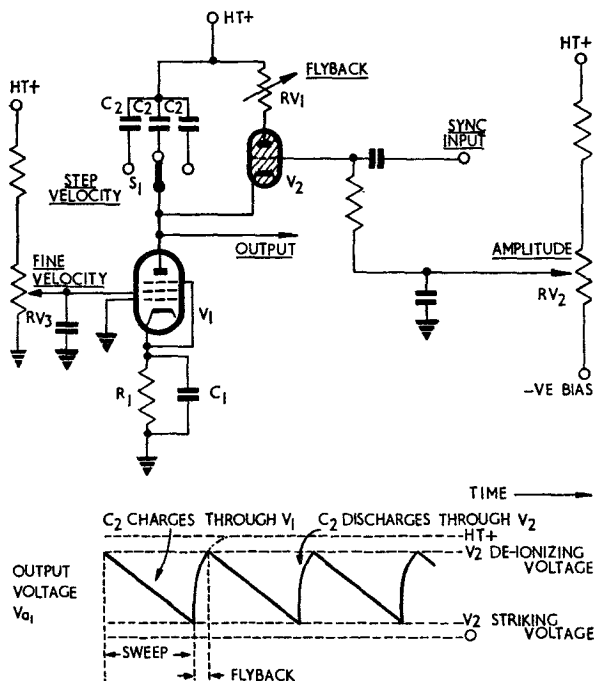


FIG 18. THYRATRON SAWTOOTH GENERATOR, CONSTANT CURRENT CHARGING

Fig 18 shows how this principle is applied in a practical circuit. It is essentially similar to the circuit of Fig 13a with the pentode V_1 replacing the charging resistor R_1 .

On applying h.t., the capacitor C_2 charges linearly and the output voltage V_{a1} falls linearly until the voltage across C_2 is sufficient to cause the thyatron V_2 to strike. C_2 is then discharged *rapidly* by V_2 , and V_{a1} rises exponentially to h.t.+. With the discharge of C_2 the voltage across V_2 falls rapidly causing V_2 to cut off. C_2 again commences to charge linearly through V_1 and the action is repeated automatically. The circuit is thus *free-running*.

The following controls are provided:

- a. The switch S_1 provides a 'step velocity control' by switching in different value capacitors for C_2 .
- b. RV_1 provides the 'flyback control' and therefore a 'fine frequency control'.
- c. RV_2 is the 'amplitude control'. It varies the bias on V_2 grid thus varying the anode voltage at which V_2 will strike.
- d. RV_3 is a 'fine velocity control' which varies the voltage applied to the *screen grid* of V_1 . Since the grid bias of V_1 is fixed by R_1C_1 , RV_3 thus controls the anode current of V_1 and hence the charging current of C_2 . The *rate* at which C_2 charges is therefore controlled.
- e. A sync pulse input may be applied to the grid of V_2 .

In a free-running timebase generator of this type the controls are *interdependent*. S_1 varies the velocity in coarse steps and also the frequency; RV_3 varies velocity and frequency; RV_2 varies both the amplitude and the frequency.

Disadvantage of Soft Valve Timebases

Because it takes time for the gas in a thyatron to de-ionize there is a limit to the rate at which the valve may be switched on and off. Since, in the timebase circuit, the thyatron is used as a switch to discharge the sawtooth capacitor, this fact limits the *frequency* of the timebase—to a maximum of about 40 kHz. Because of this, 'hard' valves are often used instead of thyatrons in sawtooth generators.

Simple Hard Valve Timebase

This circuit works in much the same way as the thyatron sawtooth generator of Fig 18. However in the hard valve circuit (Fig 19a) a 'hard' triode V_2 replaces the thyatron. Since a hard valve does not swing rapidly from conduction to cut-off and *vice versa* with a change of anode voltage (as does a thyatron) the rise and fall of capacitor voltage cannot now be used to provide the switching action. Instead, a *switching square wave* must be applied to the control grid of V_2 to cut the valve on and off. Such a circuit is therefore *no longer free-running*.

During the negative-going portion of the switching square wave, V_2 cuts off and C_2 charges through the pentode V_1 causing the output voltage to fall linearly. At the end of the sweep time, the switching waveform brings V_2 into conduction and C_2 discharges rapidly through the low resistance of the conducting triode, causing the output voltage to rise towards h.t.+. The circuit remains in this condition until the switching square wave again cuts V_2 off. The whole action of the timebase generator is thus controlled by the switching waveform. Note particularly that the *duration* of the sweep portion of the sawtooth depends upon the duration of the negative-going portion of the switching square wave.

The output from V_1 anode approximates to the required sawtooth waveform, and although the flyback is exponential this is not important as the c.r.t. is usually blanked out during this time.

The simple circuit described above is referred to as a *gated* timebase generator because the switching square wave (the gating waveform) both *starts* and *stops* the sawtooth (see Fig 19b). Gated timebase circuits are often used in radar. There are also timebase circuits in which the input pulse only *starts* the sawtooth, *ie* triggers it off. In such circuits the flyback occurs

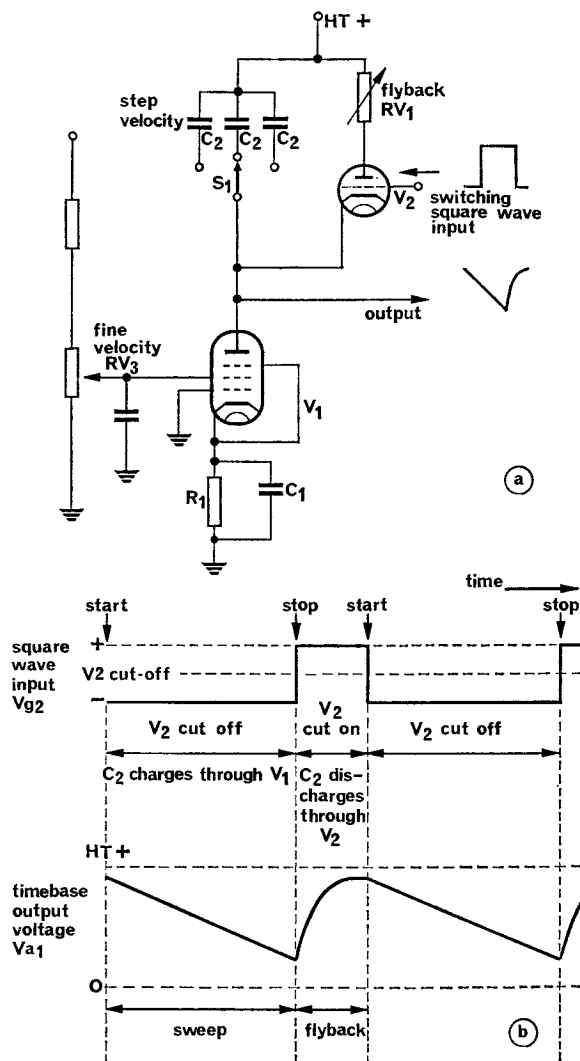


FIG 19. SIMPLE HARD VALVE TIMEBASE GENERATOR

automatically. A circuit in which the input pulse merely initiates the action is said to be *triggered*. Triggered timebase circuits are also used in radar and are considered in the next chapter.

In a gated timebase circuit, such as that of Fig 19, if the duration of the gating pulse is *constant*, the velocity controls will alter the trace *length* as well as altering the velocity. This happens because a steeper slope to the sawtooth (increased velocity) means that the fall of voltage will be *greater* during the time that the switching valve is cut off, *ie* the sawtooth *amplitude* will be greater (Fig 20).

To maintain the same trace length for different timebase velocities, the *duration* of the gating pulse has to be altered (Fig 21).

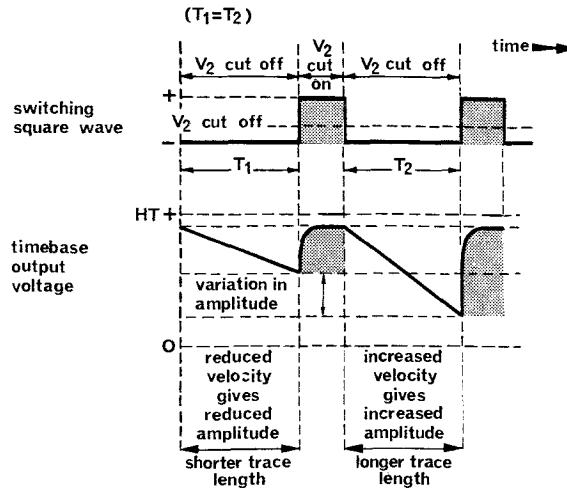


FIG 20. EFFECT OF VELOCITY CONTROLS, CONSTANT DURATION GATING WAVEFORM

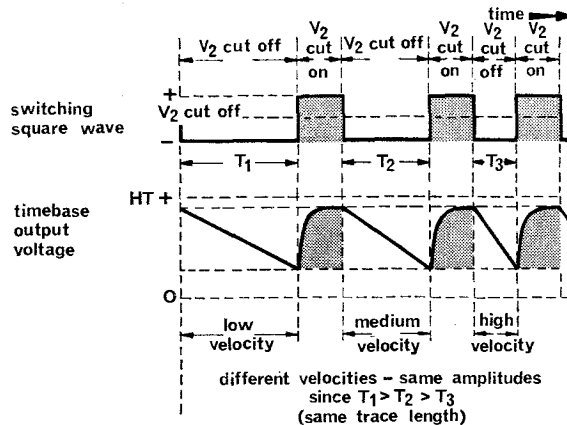


FIG 21. VARIATION OF GATING PULSE DURATION

The triggered timebase generators which we shall discuss in the next chapter do not have this disadvantage since they stop automatically. The sawtooth generated is thus independent of the input pulse duration.

If the gating waveform is obtained from a multivibrator, the timebase circuit produces sawtooth waveforms continuously. The combination of multivibrator and hard valve sawtooth generator then forms a 'free-running' timebase generator. In one arrangement, known as the *Puckle* timebase generator, the gating valve V_2 acts as one valve of the multivibrator in addition to its normal function. The circuit is illustrated in Fig 22. Examination of Fig 22 will show that the circuit is essentially the same as that of Fig 19, with the addition of the stage V_3 . V_1 and V_2 perform the same function as before, whilst V_2 and V_3 together act as a multivibrator (which can, if necessary, be synchronized—see later).

During the time that C_2 is charging, V_2 is kept cut off by the bias between its grid and cathode: V_2 cathode starts at h.t. + and falls linearly as C_2 charges; V_2 grid is connected to V_3 anode and, whilst V_3 is conducting, V_2 grid voltage is at a low value. A point is eventually reached (point X in Fig 22b) where V_2 cathode voltage falls sufficiently to cut V_2 on and flyback commences. The amplitude control RV_2 sets the 'bias' on V_2 grid and so determines the point at which V_2 cuts on.

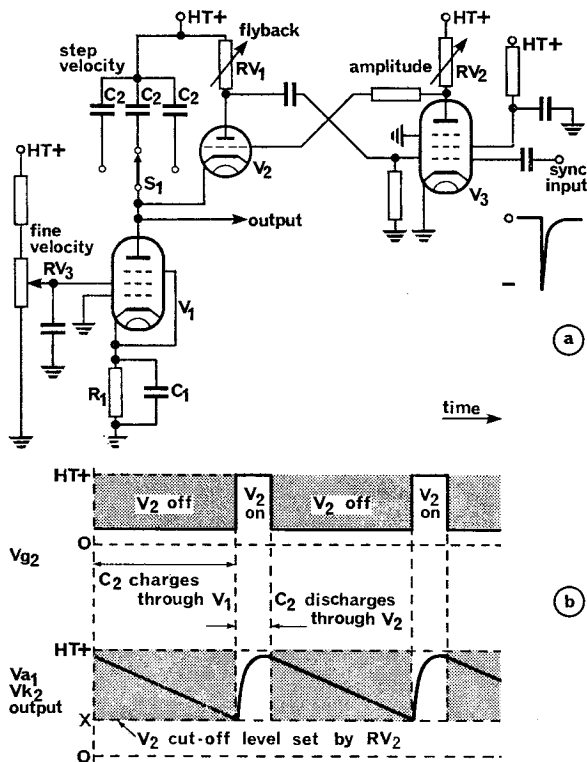


FIG 22. PUCKLE TIMEBASE GENERATOR, CIRCUIT AND WAVEFORMS

This circuit may also be synchronized by applying *negative-going* sync pulses to V_3 grid. This causes V_3 anode voltage (and also V_2 grid voltage) to rise and initiate the multivibrator action earlier than normal.

The Puckle timebase generator provides a suitable output voltage when timebase velocities are required to be continuously variable. For this reason it is widely used in oscilloscopes.

Radar Timebase Requirements

In oscilloscopes, we have seen that the usual requirement is for a free-running or synchronized timebase generator so that a wide range of different types and shapes of input may be examined. In radar however we usually need gated or triggered timebases because the timebase must be 'tied' to the p.r.f. of the transmitted pulses in such a way that the spot starts its movement along the trace at the instant the transmitter fires each pulse. The speed at which the spot then moves depends upon the maximum range to be measured and the timebase velocity is adjusted for each range to ensure that the end of the trace coincides with maximum range. The spot then flies back to the start ready for the next trace to commence coincident with the next transmitter pulse.

Fig 23 shows the block schematic diagram of the circuits which might be concerned with the production of the timebase in a typical radar equipment.

The output from the master timing unit triggers the modulator and so controls the transmitter action. The timing pulse is also used to trigger a flip-flop in the indicator so that the flip-flop action commences at the instant each transmitter pulse is fired. The output from the flip-flop is, in turn, used to control the action of the timebase generator to ensure that the sweep

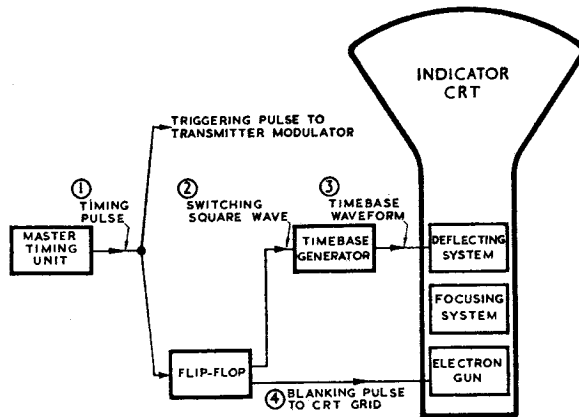


FIG 23. TYPICAL RADAR TIMEBASE ARRANGEMENTS

action is switched on coincident with each transmitter pulse. The flip-flop also provides the blanking pulse to cut off the c.r.t. during the flyback.

We have seen that if the maximum range of the equipment is 80 miles, the sweep time must be $860 \mu\text{s}$. We therefore adjust the pulse duration control in the flip-flop to make the pulse duration of the gating waveform exactly $860 \mu\text{s}$ (Fig 24). The gating waveform allows the sweep action to continue for $860 \mu\text{s}$ and then switches it off.

If the p.r.f. of the timing pulses from the master timing unit is 1,000 p.p.s., the interval between transmitted pulses is $1,000 \mu\text{s}$. As each sweep in our example lasts for $860 \mu\text{s}$, the flyback time is $1,000 - 860 = 140 \mu\text{s}$. The required waveforms are illustrated in Fig 24.

Remember that in this circuit if we wish to change ranges, the *timebase velocity* must be changed and so must the *duration* of the *gating* pulses.

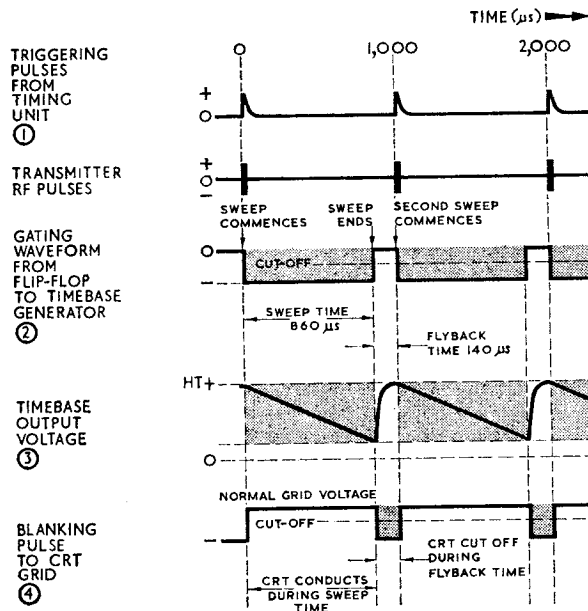


FIG 24. TYPICAL RADAR TIMEBASE WAVEFORMS

MILLER TIMEBASE CIRCUITS

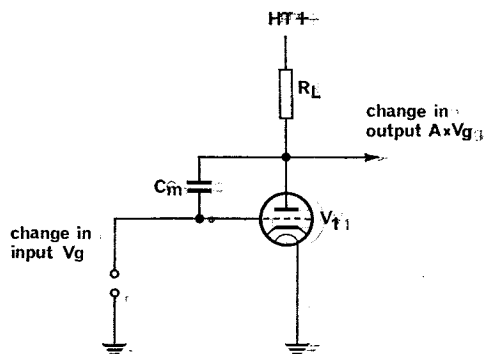
Introduction

The main limitation of the timebase generators considered in the previous chapter is that none of them can combine good linearity with a large amplitude output. To provide a large amplitude output, a large sawtooth capacitor is required. However, this also increases the flyback time and so limits the frequency of operation of the timebase. One circuit which overcomes this disadvantage is the *Miller* timebase generator. The Miller circuit provides a large effective capacitance during the sweep, but during the flyback the capacitance is much smaller. This is due to the *Miller* effect.

Miller Effect

The miller effect in amplifiers was discussed in AP 3302 Pt 1B but we now want to consider the effects of an *actual* capacitor connected between the grid and anode of a valve amplifier (Fig 1). The effects of the interelectrode capacitances can be ignored as they are very small in size compared to C_m . When a fall of voltage V_g is applied to the grid the anode rises by a voltage $A \times V_g$, where A is the amplification of the stage. The voltage across the capacitor C_m has therefore changed by $V_g + AV_g$, ie $V_g(1 + A)$, as the grid and anode changes are in antiphase.

The change of charge ($Q = CV$) on the capacitor is therefore $C_m \times V_g \times (1 + A)$. In order to change the charge across C_m current must flow through the capacitor and this can only be drawn from the input. As far as the input is concerned, a voltage change V_g has resulted in a change of charge of $C_m \times V_g \times (1 + A)$ so it is seeing an effective input capacitance of $C_m \times (1 + A)$. Let us see how this effect is used in Miller timebase circuits.



effective input capacitance will be $C_m(1+A)$ when V_t is acting as an amplifier.

FIG 1. MILLER EFFECT

Basic Miller Timebase Generator

Fig 2a shows the circuit of the basic Miller timebase generator. The most important feature of this circuit is the capacitor C_g connected between anode and grid. In addition, the load R_L is large enough to cause *bottoming* at a low value of anode voltage. The waveforms obtained (Fig 2b) are essentially due to two facts:

- During the time that the valve is working as an amplifier and Miller feedback is effective, the input capacitance has an effective value of $C_g(1 + A)$ and is discharging on a time constant of $C_g(1 + A)R_g$ seconds during the sweep time.
- When the valve is cut off and ceases to work as an amplifier, Miller feedback ceases. The effective value of C_{IN} then falls to its actual value of C_g and the capacitor is charging on a time constant of C_gR_L seconds during the flyback time.

The basic Miller circuit is a *gated* timebase generator and, until the c.r.t. trace is required, the circuit is held in a steady state by the gating waveform applied to the *suppressor* grid. When

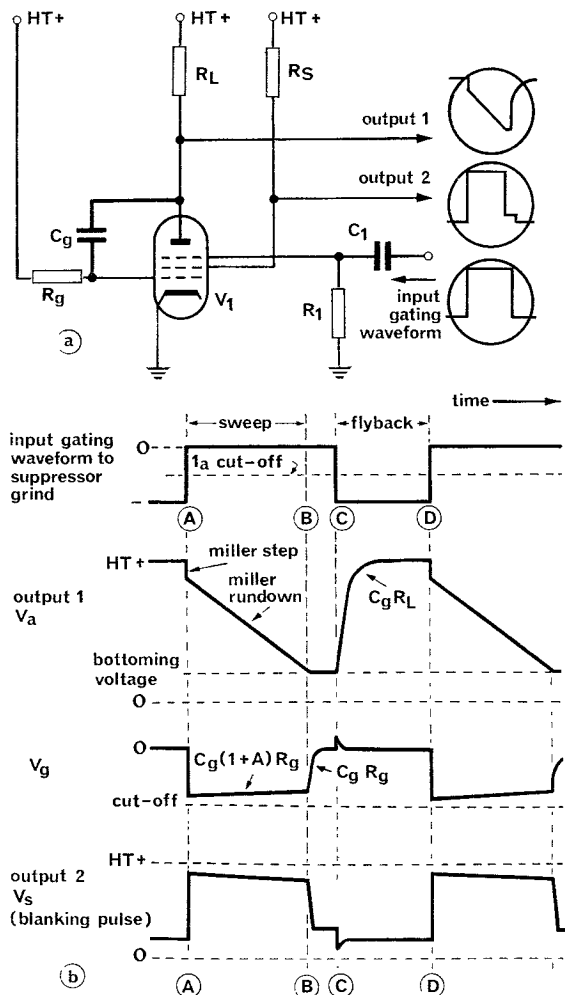


FIG 2. BASIC MILLER TIMEBASE GENERATOR, CIRCUIT AND WAVEFORMS

this is below the anode current cut-off level there is no anode current and the output V_a is therefore at h.t.+. The control grid voltage V_g is clamped to just above zero volts by the flow of grid current through R_g , and the capacitor C_g connected between anode and grid is therefore charged to the h.t. voltage. With V_g at zero volts the total space current is large and as this all flows to the screen, the screen voltage V_s is at a low value.

Instant A. When the trace is required, the gating waveform lifts the suppressor grid above the anode current cut-off level and anode current starts to flow, causing V_a to fall. A capacitor cannot change its charge instantaneously so the fall of V_a is transferred through C_g to the grid, causing V_g to fall. The initial fall in V_a (the Miller step) cannot exceed the grid base of the valve, and the fall of V_a is such that V_g is driven down almost to cut-off. It cannot fall below this point because if V_g went below cut-off, I_a would cease and V_a would rise again, lifting V_g back above cut-off, causing V_a to fall again and so on. An equilibrium is therefore reached and the fall of V_a ceases when V_g is just above cut-off. Since the valve has now almost cut itself off, the total

space current is very small and, with most of what there is going to the anode, the screen current is low so that V_s rises immediately almost to h.t. +.

Interval A to B. V_g now commences to rise as C_g discharges through R_g on a *very long* time constant of $C_g (1 + A) R_g$ seconds. Since the grid is aiming for h.t. +, the rise of V_g on this time constant represents a *very small part* of an exponential rise. V_g therefore rises *slowly and linearly*. Another way of looking at this is as follows: as V_g rises, V_a falls and the fall of V_a is transferred through C_g back to the grid to counteract the rise of V_g (the *Miller feedback*). The effect of the Miller feedback is to slow down and linearize the rise of V_g . The variation of V_g causes V_a to fall in a slow and linear manner, giving the *Miller rundown*. Amplification has of course taken place so that the fall of V_a is greater than the rise of V_g . Because V_g is rising the screen current rises and V_s falls slightly during the rundown.

Interval B to C. At B V_a reaches its bottoming voltage and with V_a held at this value *there is no further Miller feedback* through C_g to the grid so that V_g now rises *quickly* on its normal time constant of $C_g R_g$ seconds towards h.t. +. V_g is caught and held at zero volts by grid current limiting through R_g . The rise in V_g produces a larger space current and since I_a cannot increase with the valve bottomed, I_s rises and V_s falls. The circuit is now stable.

Interval C to D. At C the gating waveform to the suppressor grid falls below the anode current cut-off level and anode current ceases. V_a therefore rises to h.t. + as C_g recharges on a time constant of $C_g R_L$ seconds. The rise of V_a also lifts V_g above zero volts but this is quickly clamped to zero by the flow of grid current through R_g . The pip in the V_g waveform is reflected also in the V_s waveform. The circuit is again stable and will remain in this condition until the gating waveform again lifts the suppressor grid above the anode current cut-off level at D. The action is then repeated as from A.

The output from the anode is a negative-going timebase waveform with a *very linear* forward sweep. A square wave output can be taken from the screen grid and, after suitable d.c. restoration or clamping, can be used as the c.r.t. blanking pulse.

Although the anode voltage rundown is suitable for c.r.t. spot deflection, the spot would remain *stationary* during the time the valve was bottomed (interval B to C). For this reason the duration of the gating pulse is normally such that the anode current is cut off *before* the anode voltage bottoms. The resulting waveforms are then as shown in Fig 3.

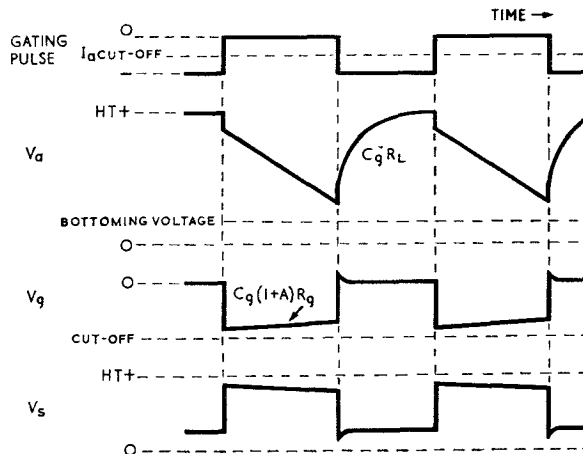


FIG 3. MILLER TIMEBASE WAVEFORMS, GATING PULSE DURATION EQUALS SWEEP TIME

Reduction of Flyback Time

The non-linearity of the flyback in a Miller timebase circuit is not usually important because the trace is blanked out during the flyback. But the fact that the flyback requires a time period of $5 C_g R_L$ seconds sets a limit on the trace recurrence frequency. In practice the flyback time may be too long. It may be reduced by using an 'anode-catching diode' as explained in p 139. In Fig 4 the starting point of the run-down is determined by the voltage on the cathode of the diode D. V_{a1} cannot rise to a voltage higher than X because as soon as it tends to do so D conducts to clamp V_{a1} to X volts. When the flyback starts, C_g is charging through R_L towards an aiming voltage of h.t. + and V_{a1} rises on a time constant of $C_g R_L$ seconds. However when V_{a1} reaches X volts D conducts and prevents V_{a1} rising further. The flyback time has therefore been reduced.

Miller Timebase Controls

Amplitude controls. It may be seen from Fig 4 that if the voltage X at the cathode of the diode were varied, this would provide an effective amplitude control for altering the trace length on the c.r.t. The arrangement is shown in Fig 5a. The lower the setting of R_{V1} , the lower is the voltage X and the smaller is the rundown before the bottoming voltage is reached. Note however that R_{V1} does not affect the slope of the rundown, i.e. the rate of rundown or the *timebase velocity* is not affected.

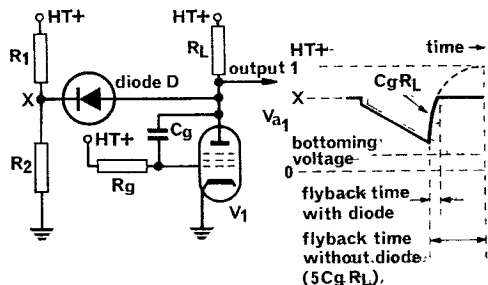


FIG 4. REDUCTION OF FLYBACK TIME BY USE OF ANODE-CATCHING DIODE.

using a limiting diode D_2 . With the diode connected as shown, V_{a1} can fall only until the voltage across D_2 is such that D_2 will conduct. This level is set by R_{V2} .

Note that the wipers of R_{V1} and R_{V2} are decoupled to earth through capacitors. This ensures that the wiper potential, once set, does not alter when the diode conducts and passes extra current through the potentiometer.

Velocity control. The rate at which V_{a1} falls in a Miller timebase circuit is controlled by the rate at which the grid voltage V_g rises. During the Miller rundown, V_g rises at a rate depending mainly upon the values of C_g , R_g and the aiming voltage to which R_g is returned; an increase in the value of either C_g or R_g slows down the rate of rise of V_g ; an increase in the aiming voltage speeds up the rate of rise of V_g . Variation of any of these factors will vary the rate of rundown (i.e. the *timebase velocity*). The factor usually varied in practice is the *aiming voltage*. This may be done by means of the 'velocity control' R_{V3} shown in Fig 6 opposite.

A complete Miller timebase circuit incorporating anode-catching diodes and also the controls mentioned above is shown in Fig 7 opposite. The action of the circuit is as previously described and R_{V1} , R_{V2} and R_{V3} are the controls dealt with above. Remember that R_{V1} and R_{V2} both affect the amplitude of the output waveform and, thus, the trace length; R_{V3} is the velocity control and affects the *time duration* of the trace (i.e. the *range* in radar). The diode D_3 connected to V_1 suppressor grid is a normal clamping diode which prevents the suppressor grid rising above zero volts.

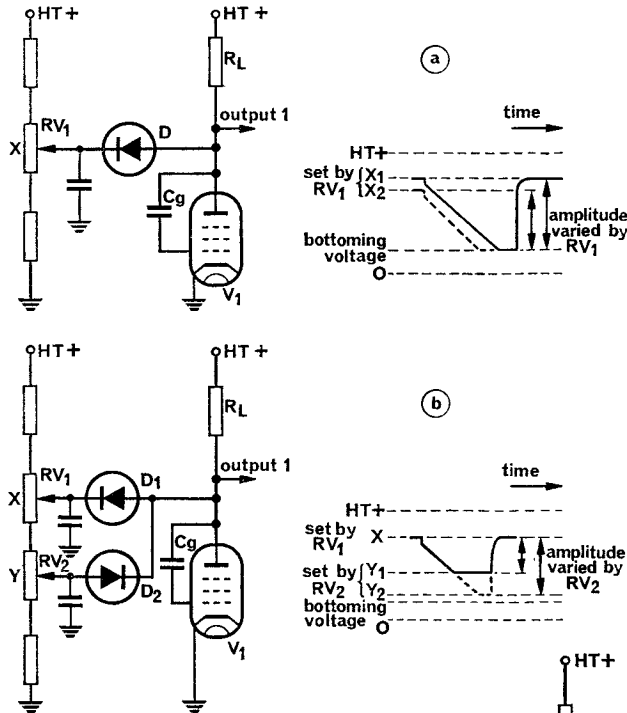


FIG 5. VARIATION OF AMPLITUDE IN MILLER CIRCUIT

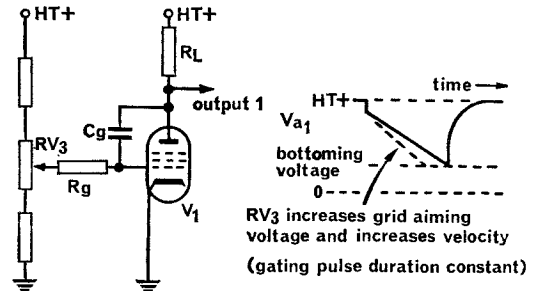


FIG 6. VARIATION OF TIMEBASE VELOCITY IN MILLER CIRCUIT

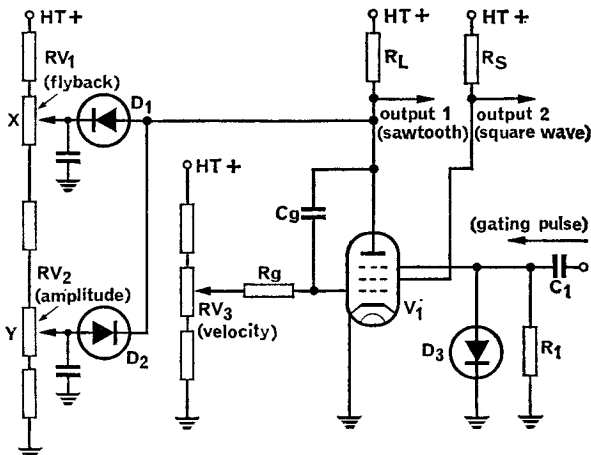


FIG 7. COMPLETE MILLER TIMEBASE GENERATOR WITH CONTROLS

Miller-transitron Timebase Generator

The main disadvantage of the Miller circuit is that a *gating waveform* is needed and for accurate operation the pulse duration of the gating waveform must be very carefully controlled. More satisfactory results may be obtained if the timebase generator itself could generate the gating waveform (*ie* if the circuit were 'self-gating'). This is done in the Miller-transitron circuit shown in Fig 8a. This single-valve, self-gating circuit performs two functions: one part of the

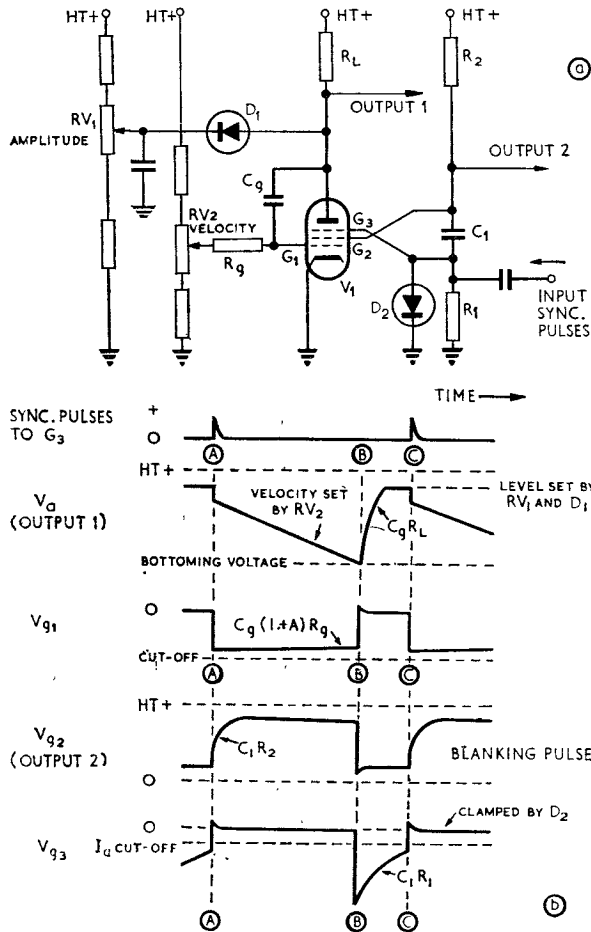


FIG 8. MILLER-TRANSITRON TIMEBASE GENERATOR, CIRCUIT AND WAVEFORMS

circuit operates as a transitron square wave generator (see p 136), the square wave output being used as the gating pulse to the other part of the circuit which operates as a Miller timebase generator.

The circuit may be free-running, synchronized or triggered. In radar, the requirement is for a close tie-up between the p.r.f. of the installation and the sweep of the timebase. Thus synchronized or triggered systems are needed. Fig 8a illustrates a synchronized circuit.

Let us suppose that the circuit is operating and that the *suppressor* grid voltage V_{g3} is below the anode current cut-off level but rising towards zero volts on a time constant of $C_1 R_1$ seconds.

to rise further. The transitron action is *cumulative* and results in V_{g2} being driven well below I_a cut-off. I_a falls to zero and V_a rises. V_{g1} is clamped to zero volts by grid current limiting and V_{g2} falls to a low value.

Interval B to C. The circuit is now stable whilst V_a rises towards h.t. + as C_g recharges through R_L and whilst V_{g3} rises towards zero as C_1 discharges through R_1 . At C the next sync pulse lifts V_{g3} above I_a cut-off and the action repeats.

Other Self-gating Timebase Generators

Many other circuits arrangements may be used as self-gating timebase generators. We shall mention only two at this stage—the *sanatron* and the *phantastron*. The action of both circuits is considered earlier in these notes (see p 137). There the emphasis was on the use of these circuits as square wave generators but it was stated that a sawtooth timebase output may also be provided. The action of these circuits will not be repeated here. It is sufficient to know that in each case one part of the circuit acts as a flip-flop to produce a gating square wave from a trigger input, whilst the other part of each circuit acts as a gated Miller timebase generator. The circuits and waveforms are shown for convenience in Fig 9. One point previously mentioned is worth repeating—the Miller step in the phantastron is larger than the usual Miller step because of the cathode follower action of this circuit. We shall see later how use can be made of this fact.

Transistorized Gated Miller Timebase Circuit

One form of gated Miller timebase generator using transistors is illustrated in Fig 10. In this circuit a bias voltage of $-3V$ is applied via the diode D to the *emitter* of TR_2 . Thus when the gating waveform applied to TR_2 base is at zero volts, TR_2 is cut off and the collector of TR_2 is at $-12V$. When the gating waveform rises to $-5V$, TR_2 starts to conduct and the Miller step is produced. TR_2 emitter then rises to approximately the same voltage as its base and D cuts off. This means that the current through TR_2 is now controlled by TR_1 which, in turn, is being controlled by the *feedback* from TR_2 collector via the Miller capacitor C. Because of the Miller feedback, the current flowing through R_L from TR_1 and TR_2 rises *linearly* as C discharges through R. The collector voltage of TR_2 thus *falls linearly* until it equals the gating voltage. The transistor has now bottomed and TR_2 collector voltage remains at the bottoming voltage until the gating waveform returns to zero volts, cutting off TR_2 . Flyback then occurs, TR_2 collector voltage returning to $-12V$ as C recharges through R_L .

Transistorized Self-gating Miller Timebase Generator

The circuit of a *triggered* timebase circuit which produces its own gating pulse is illustrated in Fig 11. In this circuit TR_1 and TR_2 perform the same function as in Fig 10. TR_3 provides the *self-gating* action. In the stable condition, before the application of a trigger pulse, TR_1 and TR_3 are conducting. The collector voltage of TR_1 is approximately equal to the base voltage of TR_3 (ie $-1.5V$) and TR_3 collector is at $-3V$ (via D_1). Diodes D_1 and D_2 are both conducting so that the base and emitter of TR_2 are both at the same voltage ($-3V$) and TR_2 is cut off. TR_2 collector is thus at $-12V$.

When a *negative-going* trigger pulse is applied to the *emitter* of TR_3 , this transistor cuts off and its collector voltage rises to $-6V$. This voltage, applied to the base of TR_2 , switches TR_2 on and TR_2 conducts, producing the Miller step. The current through TR_2 is controlled by TR_1 which itself is being controlled by the feedback from TR_2 collector via the Miller capacitor C. Because of this negative feedback the current in TR_1 falls to a low value and TR_1 collector voltage rises to about $-2V$, sufficient to hold TR_3 cut off.

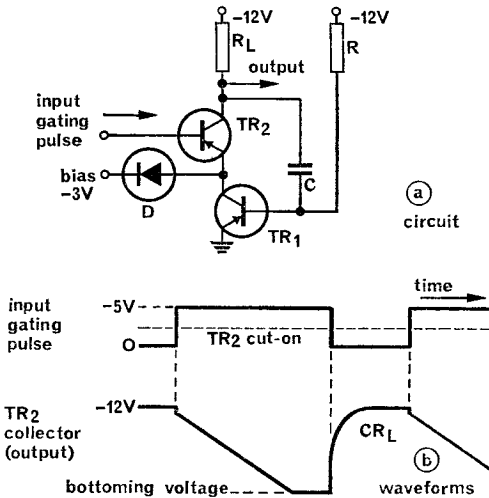


FIG 10. GATED MILLER TIMEBASE CIRCUIT USING TRANSISTORS

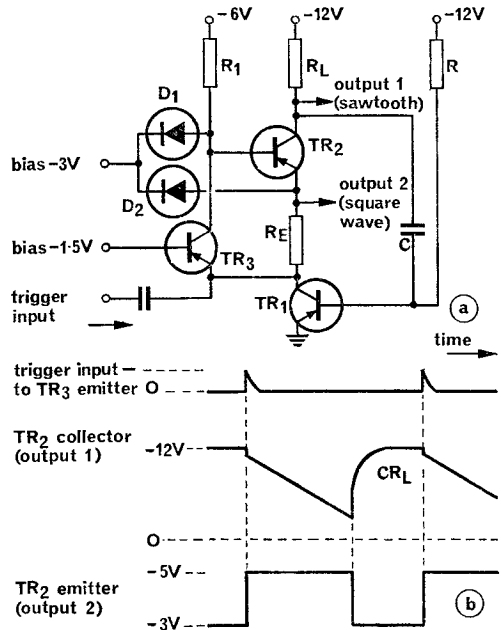


FIG 11. SELF-GATING MILLER TIMEBASE CIRCUIT USING TRANSISTORS

Miller rundown now occurs. The current flowing through TR_1 and TR_2 increases linearly, causing TR_2 collector voltage to *fall linearly* from $-12V$ as C discharges through R . This fall continues until TR_1 collector voltage (and TR_3 emitter voltage) rises sufficiently to cut TR_3 on.

When TR_3 cuts on its collector voltage falls to $-3V$ and this causes TR_2 to cut off. The collector voltage of TR_2 therefore flies back to $-12V$ as C recharges through R_L . The circuit remains in this condition until another trigger pulse is received.

The sawtooth timebase waveform is taken from the collector of TR_2 . An output may also be taken from the emitter of TR_2 : this is a *square wave* output whose duration corresponds to the sweep time and, after suitable restoration, may be used as a blanking pulse to eliminate the effect of the flyback on the screen.

Unijunction Timebase Generator

The circuit shown in Fig 12 act as a simple sawtooth generator. The basic action of the unijunction transistor is explained earlier in these notes (see p 152). In this circuit, TR_1 is a p-n-p transistor which acts as a *constant current* source for *linear* charging of C (a transistor has characteristics similar to the I_a-V_a curve of a *pentode*). On connecting the supply, C commences to charge linearly through TR_1 , causing the emitter voltage of the unijunction (and also the output voltage) to rise linearly. When the unijunction emitter voltage is sufficiently positive to overcome the reverse bias between it and $base_1$, the unijunction 'fires' and C discharges rapidly through the forward-biased emitter-base₁ junction and R_1 . When the unijunction stops conducting, C recharges and the action is repeated. This circuit is therefore that of a *free-running* timebase generator whose frequency is determined by the time constant formed by C and TR_1 and also by the voltage at which the unijunction fires.

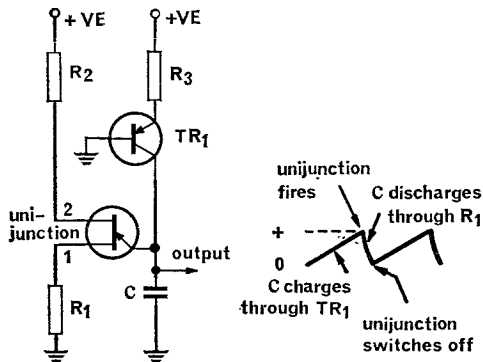


FIG 12. UNI-JUNCTION SAWTOOTH TIMEBASE GENERATOR

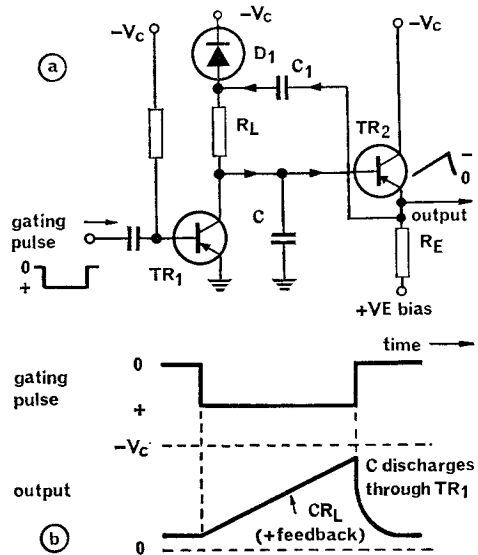


FIG 13. GATED BOOTSTRAP TIMEBASE GENERATOR, CIRCUIT AND WAVEFORMS

The Gated Bootstrap Timebase Generator

The basic 'bootstrap' timebase circuit is shown in Fig 13. In the quiescent state, TR_1 is conducting and the sawtooth capacitor C is discharged. The output voltage, from the emitter-follower TR_2 , is then almost zero. On receipt of a large-amplitude positive-going gating pulse, TR_1 cuts off and the capacitor C commences to charge via R_L and the diode D_1 . This negative-going voltage waveform is fed to the base of the emitter-follower TR_2 , causing the emitter voltage (and also the output voltage) to rise. This rise in voltage across R_E is fed back via C_1 to the top end of R_L . If TR_2 is acting as a perfect emitter-follower and R_L is a very high value, the rise in voltage at the top end of R_L equals the rise in voltage at the bottom end of R_L , thus keeping the voltage across R_L constant. The current through R_L is therefore maintained constant and since this is the charging current the sawtooth capacitor, C charges linearly. The output voltage during the sweep will therefore rise in a linear manner. Without the diode D_1 the top end of R_L could not go more negative than $-V_C$. When the bottom end of R_L goes more negative the top end of R_L must vary in a similar manner to keep the voltage across R_L constant. If this variation is such that the top end of R_L goes more negative than $-V_C$ then the diode D_1 cuts off to isolate V_C supply point from R_L . This allows the follower action across R_L to be maintained to ensure linear charging of C .

In practice the gain of an emitter-follower is slightly less than unity so that the rising voltage fed back to the top end of R_L via C_1 is not quite equal to the rising voltage at the bottom end of R_L . The result is that C charges slightly non-linearly and the output during the sweep has a slight exponential curve. Certain modifications may be made to the basic circuit to improve the linearity. These are not considered here.

At the end of the sweep time the gating waveform returns TR_1 to conduction and C discharges through it during the flyback. The waveforms are illustrated in Fig 13b. Note that there is no Miller step in the sawtooth waveform.

OTHER RADAR TIMEBASE GENERATORS

Introduction

So far we have considered the sawtooth voltage waveforms and the timebase generators required to produce spot deflection in *electrostatic* c.r.t.s. Although electrostatic c.r.t.s are used in most oscilloscopes, in radar they are usually confined to deflection-modulated displays such as the type A. The magnetic c.r.t. is more usual for other radar displays. There are two main reasons for this choice:

- a. We can produce a higher beam current in a magnetic c.r.t. because the beam is not confined by anode structures or deflector plates as in the electrostatic type. In addition we can usually apply higher values of e.h.t. to a magnetic c.r.t. The result is a much brighter picture which is ideal for intensity-modulated displays. (A television picture is a good example of this bright display.)
- b. Magnetic deflection is usually more suitable for producing radial timebases such as the p.p.i.

The main disadvantages of the magnetic c.r.t. are the greater difficulty in producing the required timebase and the power dissipated in the focusing and deflecting coils. These are the main reasons for not using a magnetic c.r.t., for example, in an oscilloscope.

In this chapter we are going to discuss how displays *other* than the type A are produced. To do this we must first examine timebase generators for magnetic c.r.t.s.

Timebase Generator for Magnetic CRT

In a magnetic c.r.t. the spot is deflected by magnetic fields which are produced by *currents* flowing through the deflecting coils. To produce a horizontal trace on the screen of the c.r.t. the timebase voltage is applied to the X deflecting coils as in Fig 1.

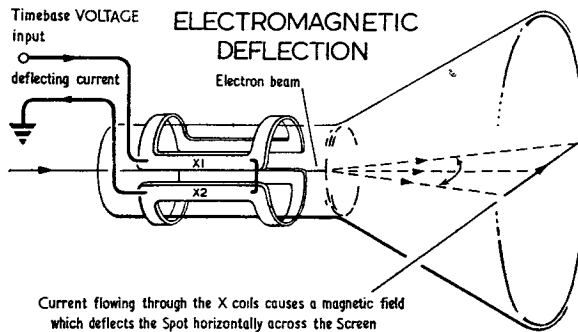


FIG 1. DEFLECTION IN A MAGNETIC CRT

As in all timebases, we require the spot to move at a *constant velocity* during the sweep time. To achieve this in a magnetic c.r.t. we must make the deflecting *current* (not voltage) vary in a linear manner. It may appear that if we were to apply a sawtooth voltage to the deflecting coils that the current would vary in a like manner. This is not so however because the inductance of the coils *opposes changes* in current. Fig 2a shows that the deflecting coils contain both inductance and resistance. Thus if a square wave of voltage is applied to the coils the current rises and falls *exponentially* on a time constant of $\frac{L}{R}$ seconds. If we apply a timebase voltage

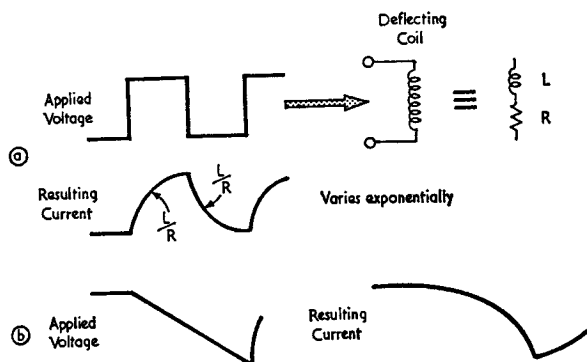


FIG. 2. NON-LINEAR VARIATION OF CURRENT IN DEFLECTING COILS

which changes linearly during the sweep time, the waveform of deflecting current will be as shown in Fig 2b.

To make the change of current more linear we could feed the coils from a high resistance source (e.g. a pentode). If, in addition, the circuit used a large amount of negative feedback the current change would be linear enough for many applications.

In practice however it is more usual to decide what the *shape* of the *voltage waveform* applied to the deflecting coils must be to cause the *current* through the coils to rise *linearly*, and then to use a circuit to generate the required voltage waveform. To do this it is easier to assume a linear rise of current through the coils and then add up the resulting voltage drops across the inductive and resistive parts of the coil to find the total input voltage. This is shown in Fig 3.

In Fig 3a the desired change in current is shown. This current, flowing through R , will produce a sawtooth voltage V_R across R , because a resistance obeys Ohm's Law (Fig 3b). Across L , however, the back e.m.f. V_L depends upon the *rate* at which the current is changing. At A we have a *sudden* variation in the rate of change of current so V_L rises instantly; from A to B the current is changing *at a constant rate* so V_L is constant; and at B we have another sudden variation in the rate of change of current, in the *reverse* direction to that at A, so V_L falls instantly. The voltage across L is thus a *square wave* as shown in Fig 3c. The total applied voltage V is the sum of V_R and V_L and this is shown in Fig 3d. This waveform is usually referred to as a *pedestal waveform*. Other names are *trapezoid* and *step-sawtooth* waveforms.

The normal method of feeding the deflecting coils is to place them in the anode or the cathode circuit of a power valve and to control the current through the

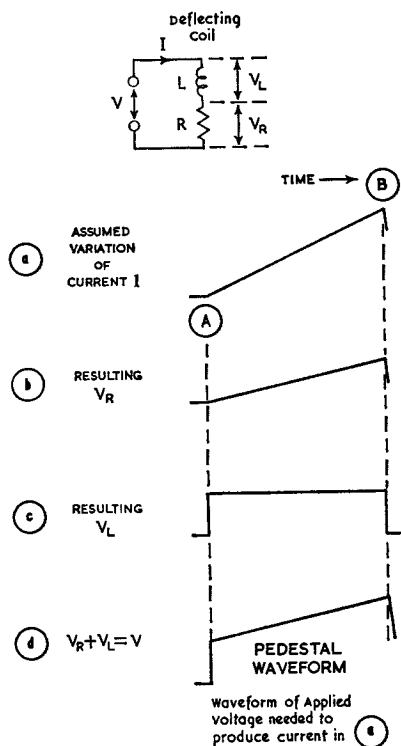


FIG. 3. THE PEDESTAL WAVEFORM

coils by variations in the grid voltage V_g (Fig 4). To provide a linear current sawtooth, V_g must be a *pedestal* waveform. We shall now see how this waveform is produced.

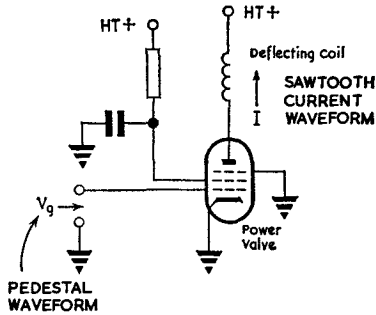


FIG 4. METHOD OF FEEDING DEFLECTING COILS

Pedestal Waveform Generators

The pedestal waveform may be produced by one of several circuits. Examination of the waveform shows that it is similar in many ways to the output from a Miller timebase generator. The only real difference is that the sweep portion of the pedestal waveform commences with a *very large step*. We have already dealt with a Miller-type circuit which produces a step bigger than normal, *ie* the *phantastron* (see p 197). Thus it is possible to use a phantastron to produce the required pedestal waveform.

Another arrangement which is used in some radar circuits is shown in Fig 5. This is a normal Miller timebase circuit in which the anode load consists of *two* resistors R_1 and R_2 in series. The Miller capacitor C_g is connected from the *junction* of the two resistors to the grid.

Whilst the pentode is cut off, the voltage at B (and at A also) is at h.t. + as there is no current flowing through the anode loads. When the pentode is cut on at the suppressor grid by the gating pulse anode current starts to flow. The fall of voltage at point B is fed back through C_g to the grid of the valve as in the normal Miller step. The voltage at A has fallen considerably producing a 'step' much larger than the grid base of the valve. This waveform provides the pedestal output required. The step at point A can be made any required size by correctly choosing the relative values of the two resistors R_1 and R_2 . Split anode loads are very common in the sanatron circuit where the fall at point A can be used to cut off the second, or flip flop, valve.

Another common circuit for producing a pedestal voltage waveform is shown in basic form in Fig 6 overleaf. It is very similar to the circuit of a simple hard valve sawtooth generator (see page 184) with the addition of the resistor R_1 .

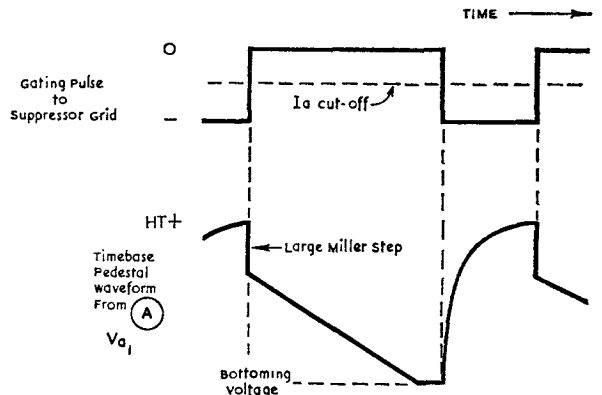
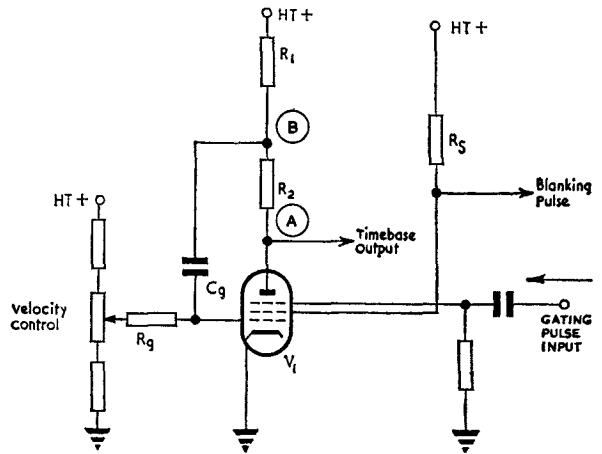


FIG 5. USE OF A SPLIT LOAD TO PRODUCE A PEDESTAL WAVEFORM

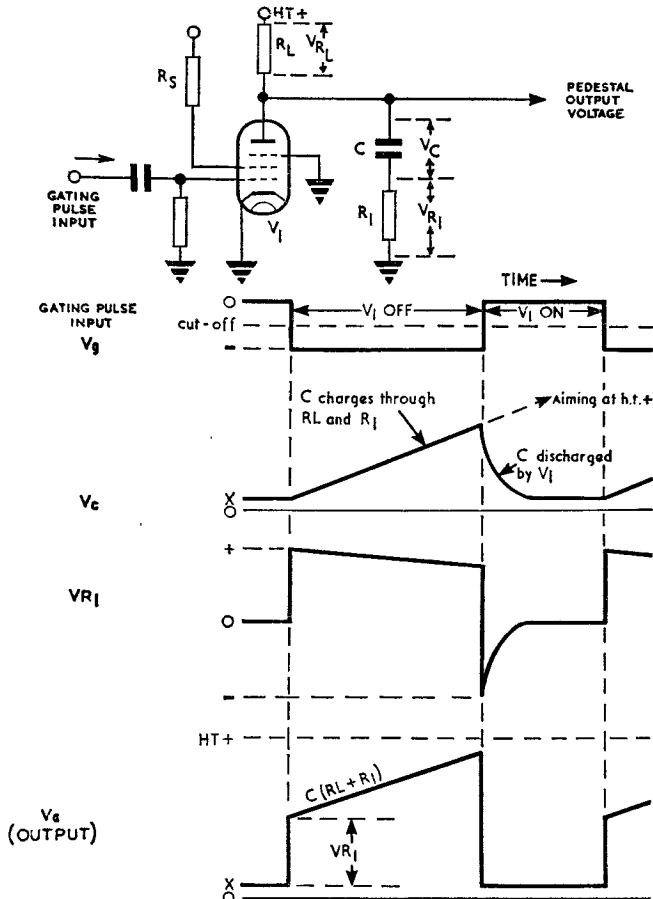


FIG 6. BASIC PEDESTAL WAVEFORM GENERATOR

The valve is switched on and off by a gating waveform. Initially, V_1 is conducting, with V_g at zero volts, and the resulting anode current produces a voltage drop across R_L which brings V_a down to a voltage $X = HT - V_{R_L}$. The capacitor C is charged to this voltage and the output is initially at this level, with V_{R_1} at zero volts.

When the gating waveform cuts off V_1 we have the full h.t. developed across R_L , C and R_1 in series. C is already charged to X volts and cannot immediately change its charge. We therefore have $(HT - X)$ volts shared between R_L and R_1 in proportion to their values and V_{R_1} immediately rises from zero. Thus the output, which is the sum of V_c and V_{R_1} , also rises immediately in a large step by the value of V_{R_1} .

C now commences to charge through R_L and R_1 towards h.t.+ on a time constant of $C(R_L + R_1)$ seconds. As it does so the output voltage rises as shown. In a practical circuit, the pulse duration of the gating waveform and the component values are such that C charges through only a very small part of its exponential and the rise in output waveform is almost linear.

When the gating waveform cuts V_1 on, C is quickly discharged through V_1 and R_1 , and the output returns to its original value of X volts. This exponential flyback is usually blanked out.

From the anode of V_1 we have a pedestal voltage waveform and this may be used as the input to the valve which is controlling the current through the deflecting coils (see Fig 4).

The magnitude of the initial step is determined by the ratio of R_L to R_1 . By decreasing the ratio $\frac{R_L}{R_1}$ whilst keeping $C(R_L + R_1)$ constant, the *step* is increased but the linear rise is unchanged.

Circular Timebases

So far we have seen how a horizontal timebase trace may be produced both in electrostatic and in magnetic c.r.t.s. There are occasions however when other forms of timebase trace are required. For example, a *circular* timebase trace is used in a type J display (see p 22).

It was mentioned in Part 1B of these notes (p 524) that if we applied sine waveforms of equal amplitude and frequency, but differing in phase by 90° , to the deflecting plates of an electrostatic c.r.t., a circular pattern as used in a type J display would result. Fig 7 shows the basic arrangement. A similar arrangement may be used for magnetic c.r.t.s.

A sine wave of voltage of the required frequency is applied to R and C connected in series, and V_C is applied to the X plates and V_R to the Y plates. V_C and V_R are 90° out of phase, their frequency is the same and, if the reactance of C equals R, their amplitude is the same. A circular timebase trace results.

In such a timebase the spot moves at a *constant speed* so that its motion represents a true timebase against which radar ranges may be measured. Compared with the type A trace a *longer* timebase is available so that for the same range coverage, targets are indicated in greater detail. Either deflection-modulation or intensity-modulation may be used for target indication.

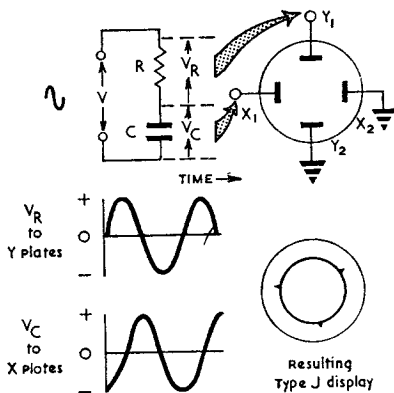


FIG 7. PRODUCTION OF CIRCULAR TIMEBASE TRACE

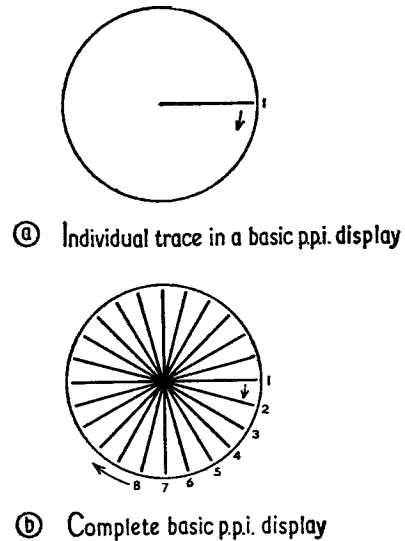


FIG 8. RADIAL TIMEBASE DISPLAY

Radial Timebases

A p.p.i. display uses a *radial* timebase. In the *basic* p.p.i. display the trace starts at the *centre* of the screen and extends to the edge (Fig 8a). This trace is obtained by applying a current sawtooth to the deflecting coils. At the same time each trace is shifted slightly round the face of the screen in relation to the preceding trace so that, with a large number of successive traces, the whole circular display area is covered (Fig 8b). As we noted earlier the trace is normally shifted round the screen in synchronism with the aerial rotation (see p 21). The result

is not unlike the spokes of a bicycle wheel but since the p.p.i. is an intensity-modulated display the 'spokes' are not usually seen.

The start of a timebase trace in radar indicates the position of the radar aerial relative to targets. Thus in the basic p.p.i. display the *centre* of the screen corresponds to the position of the radar and the display shows echoes from all points around the radar installation. Very often, objects 'behind' the radar or to one side or the other are of no interest, so that much of the display is wasted. This can be avoided by *displacing* the start of the trace from the centre of the screen. For example, the start of the trace may be 'off-centred' to the bottom of the c.r.t. screen as in Fig 8c. This allows the required portion of the display to be examined in greater detail. We shall see later how *off-centre* p.p.i. displays are produced.

Production of Radial Timebase

A radial timebase can be produced by feeding the deflecting coils with a linear current sawtooth waveform and then rotating the coils around the neck of the c.r.t. in synchronism with the aerial movement (Fig 9). This is a *rotating coil* p.p.i. display. Since the coils are rotating, the current sawtooth has to be fed to them through slip rings. This mechanical method works reasonably well for low speeds of rotation but becomes inefficient at high speeds.

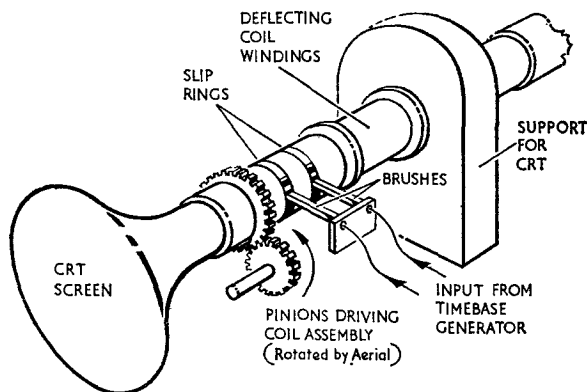


FIG 9. ROTATING COIL PPI DISPLAY SYSTEM

A much better arrangement, and the one mainly used, is the *fixed coil* p.p.i. display. In the fixed coil p.p.i. display there are *two sets* of deflecting coils, one for producing horizontal deflection of the spot and the other for vertical deflection. Sawtooth current timebase waveforms are applied to the two sets of coils simultaneously. The amplitude and polarity of each current waveform is made to vary in a regular manner, the variations in each being 90° *out of phase*. We shall see shortly how this is done. Meantime, note that the *direction* in which the resulting timebase trace lies depends upon the *relative amplitudes* of the two timebase current waveforms at any instant. This may be seen from Fig 10.

a. At time t_1 the timebase input to the X deflecting coils is zero whilst that to the Y deflecting coils has a maximum positive value. There is therefore no horizontal deflection and the timebase trace lies in position 1.

b. At time t_2 the X waveform has increased from zero in a positive direction and the Y waveform has decreased from its maximum positive value. The two waveforms have equal amplitudes, so we now have *equal* horizontal and vertical deflection and the trace lies in position 2.

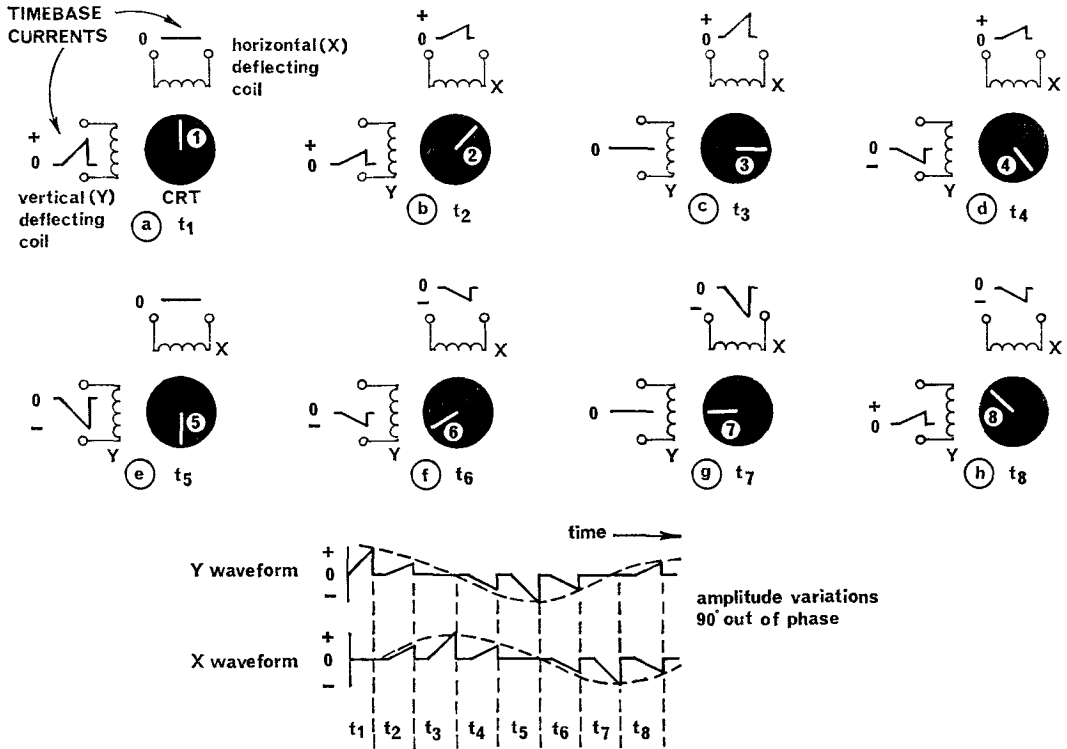


FIG 10. BASIS OF FIXED COIL PPI DISPLAY

c. At t_3 the X waveform now has its maximum positive value and the Y waveform has fallen to zero. With no vertical deflection the trace lies in position 3.

d. At t_4 the X waveform has decreased from its maximum positive value and the Y waveform has increased from zero in a *negative* direction. The resulting trace is in position 4.

The same reasoning may be applied to times t_5 , t_6 , t_7 and t_8 when the trace lies in positions 5, 6, 7 and 8 respectively. The variations in the two sets of current timebase waveforms necessary to produce this radial timebase are shown at the bottom of Fig 10. From this we can see that the two sets of current sawtooth waveforms must be amplitude-modulated by *sine waves* which are 90° out of phase, *ie* one sawtooth is modulated by a '*sine*' waveform and the other by a '*cosine*' waveform. In addition, to fill the screen area, we require many more traces at intervals of time shorter than those indicated in Fig 10. The required waveforms are therefore as shown in Fig 11.

To be of any value for indication of radar bearing the trace position must move around the screen *in step* with the rotating aerial scan. The sine and cosine modulating waves must therefore in some way be related to the position of the aerial at any instant. There are several ways of obtaining this. One of the most common is to use a *resolver synchro* (see Fig 12). The resolver synchro, mentioned in p 538 of Part 1B, is a device with two stator coils at right angles to each other, and a rotor winding which moves under the influence of the aerial.

One stator winding is connected to the X deflection coils and the other to the Y deflection coils. The rotor is energized by a pedestal voltage waveform and the resulting current in the rotor winding is a sawtooth waveform. The current produces an alternating magnetic field which links with the stator windings and induces voltage in each of them, the magnitude and polarity of the voltages depending upon the orientation of the rotor at that instant. The

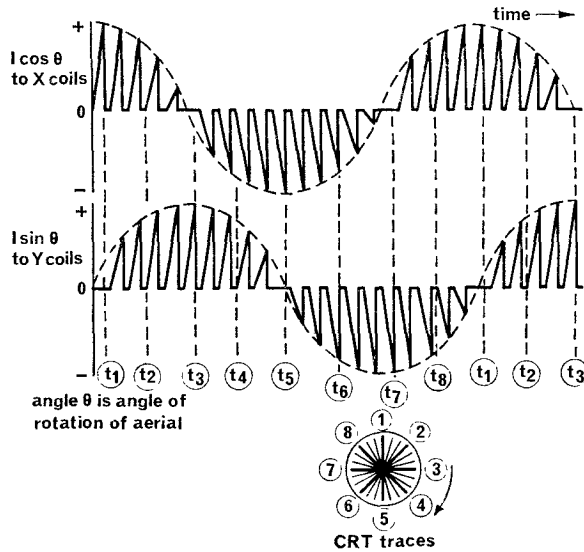


FIG 11. SINE AND COSINE MODULATED SAWTOOTH WAVEFORMS

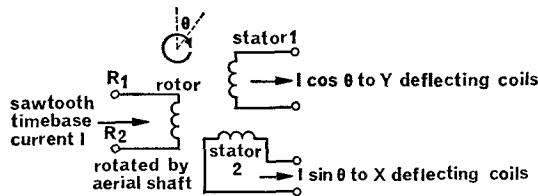


FIG 12. RESOLVER SYNCHRO SYSTEM USED IN FIXED COIL PPI DISPLAY

voltages induced in the stator windings are pedestal-shaped waveforms so that the resulting currents in the X and Y deflecting coils are the required sawtooth waveforms.

When the rotor is aligned with stator 1 the voltage induced in stator 1 has its maximum value so that the timebase current to the X deflecting coils has its maximum amplitude; the voltage induced in stator 2 at this time is zero and the Y deflecting coils have no timebase current input. If the rotor is turned through 90° these conditions are reversed: the X deflecting coils now have zero input and the Y deflecting coils have maximum amplitude timebase current.

As the rotor is turned *at a constant speed* through 360° under the influence of the aerial, the amplitude and polarity of the timebase currents fed to the X and Y deflecting coils vary sinusoidally, the variations being 90° out of phase (refer to Fig 11). In mathematical terms we say that the output from the stator windings are of the form $I \sin \theta$ and $I \cos \theta$, where I is a sawtooth current and θ is the angle through which the rotor has been turned by the aerial. These currents, when fed to the deflecting coils, produce the required radial timebase.

If we require an *off-centre* p.p.i. display, two *additional* sets of deflecting coils may be used. These two sets of coils are fed with d.c. current, the magnitude and polarity of which may be varied as required by 'shift' potentiometers. One set of coils shifts the whole display in the X axis, and the other in the Y axis. By varying the currents through these shift coils we can off-centre the display as required.

CHAPTER 15

STROBE PULSE CIRCUITS

Introduction

We have seen in earlier chapters of this book that to measure the range of a target on a type A display, a range scale may be calibrated and placed in front of the c.r.t. screen. Alternatively, a cal pip generator may be used to provide range markers along the timebase trace, each cal pip representing a given range. Both these methods suffer from the disadvantage that the number of markers we can reasonably accommodate on the display is limited. Thus when a target echo falls between two markers some guesswork is needed in estimating an accurate target range. A much better method, especially when several target echoes appear on the timebase trace, is to use a *strobe pulse* to mark a particular target.

A strobe pulse is a single pulse which is generated once every sweep period of the trace and which can be *moved along the trace* to mark any selected target. In addition, the strobe pulse is linked with the start of the timebase trace in such a way that the control responsible for the movement of the strobe pulse along the trace can be calibrated in terms of *radar range*.

In deflection-modulated displays, such as the type A, the strobe pulse is usually applied to the Y deflecting system, along with the signal input, to provide a suitable deflection on the time-base trace (Fig 1a). In p.p.i. displays the strobe pulse is usually applied to both sets of deflecting coils or plates to position the marker at any point on the face of the screen (Fig 1b); at the same time the c.r.t. blanking pulse raises the grid above cut-off so that the marker is seen.

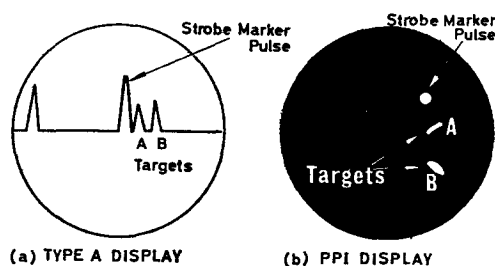


FIG 1. USE OF STROBE MARKER PULSES

Production of Strobe Pulses

Fig 2 shows how a strobe marker pulse may be produced. The input trigger pulse, which comes from the master timing unit, and which determines the p.r.f. of the whole equipment, is applied to the main timebase generator and also to a triggered square wave generator, such as a flip-flop or phantastron. The leading edge of the flip-flop output is locked to the input trigger and to the start of the main timebase, but the *trailing edge* is *variable*. The *time* at which the trailing edge is made available after the start of the main timebase is determined by the pulse duration control of the flip-flop. The flip-flop output is applied to a short CR differentiating circuit where positive- and negative-going pips of voltage are produced. The positive-going pips are removed by a limiting circuit and we are left with negative-going pips which may be used as the strobe pulses. Note that the strobe pulses are locked to the *trailing edge* of the flip-flop output so that by varying the pulse duration control of the flip-flop the strobe pulse can be moved to *any point* on the c.r.t. trace. We can also calibrate the pulse duration control in terms of *range* because the delay between the start of the trace and the generation of the strobe pulse is determined by this control. Such a control is usually referred to as a *range potentiometer*.

In Fig 2 we have shown the strobe pulse as a negative-going marker pip. Many types of strobe markers are used in practice and Fig 3 shows the ones commonly used.

Strobe Pulse Generators

Fig 2 shows only one way in which a strobe pulse may be produced. There are many other ways and in the following paragraphs we shall consider two circuit arrangements which we may meet in practice. Both circuits need *two* inputs: a d.c. voltage, which can be varied,

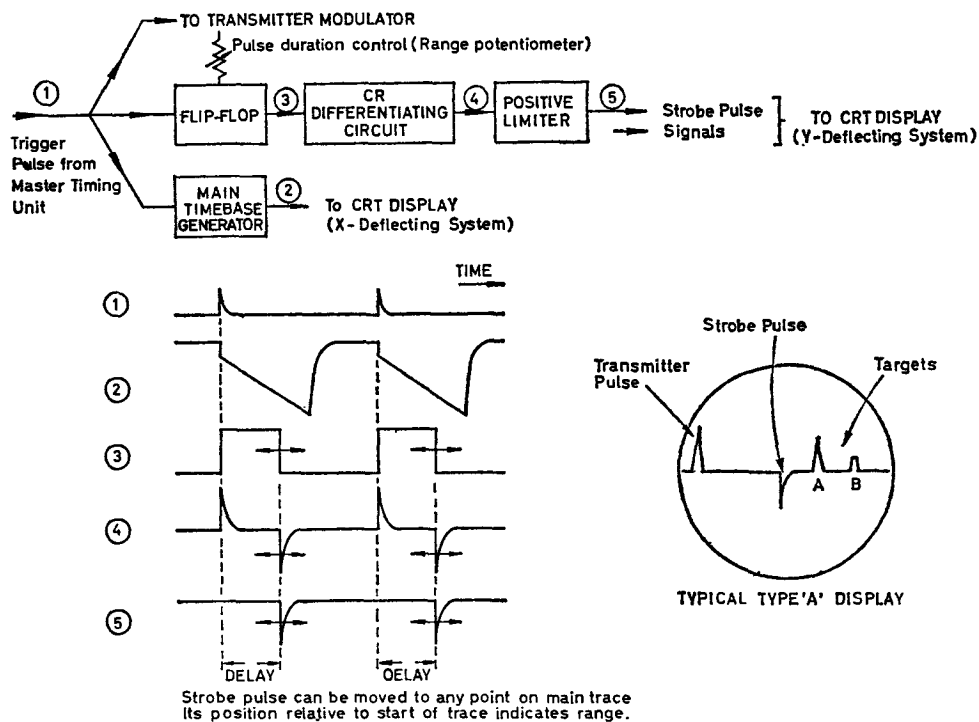


FIG 2. PRODUCTION OF STROBE MARKER PULSE

and the sawtooth timebase voltage. The circuit *compares* these two inputs and when the timebase input voltage rises (or falls) to the set level of d.c. voltage, the circuit operates to produce the strobe marker pulse.

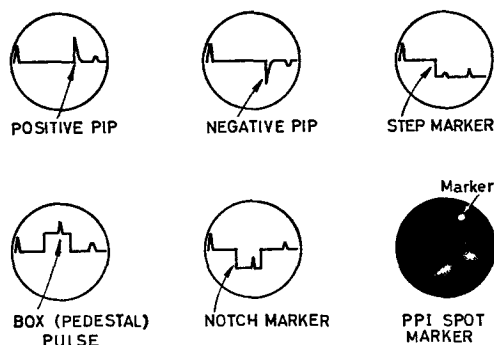


FIG 3. STROBE PULSE MARKERS IN COMMON USE

The first circuit we shall consider is the *long-tailed pair* strobe pulse generator, illustrated in Fig 4. The basic long-tailed pair circuit is mentioned briefly in p 310 of Part 1B. In Fig 4, the positive-going sawtooth timebase voltage is applied as input to the grid of V_1 ; with zero input this valve is cut off. The control grid voltage of V_2 is determined by the setting of the range potentiometer RV and, initially, this positive bias causes V_2 to conduct heavily. V_2 is however acting as a *cathode follower* so that the cathode voltage of *both* triodes is at virtually the same voltage as that of V_2 grid (determined by RV). Initially therefore, with the timebase

input zero, the grid of V_1 is at earth, its cathode is positive and V_1 is cut off. The output, taken from V_1 anode, is at h.t. +.

When the sawtooth input is applied, the grid voltage of V_1 rises. V_1 remains cut off however until the rising voltage at the grid overcomes the cathode bias set by RV. At this point V_1 conducts and V_{a1} falls. This fall is applied to V_2 grid via C_1 and V_2 conducts less heavily. By cathode-follower action the voltage at the cathode then becomes *less positive* and the anode current of V_1 increases further. V_{a1} falls further and this cumulative action rapidly causes V_1 to conduct heavily and V_2 to be cut off. This condition is maintained until the timebase voltage returns to zero at the end of the trace when the circuit reverts to its original state. V_{a1} thus falls

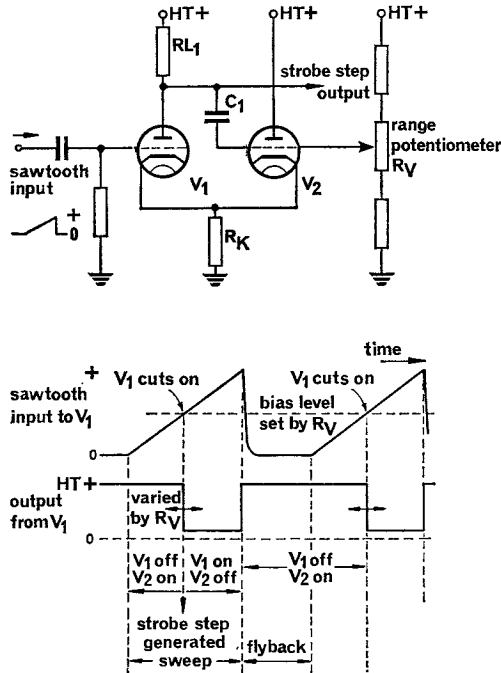


FIG 4. LONG-TAILED PAIR STROBE PULSE GENERATOR

in a *large negative-going step* and this voltage may be used as a strobe pulse step marker. Alternatively, the output from V_1 may be differentiated to produce a negative-going strobe marker pip. The point on the timebase at which the strobe appears is controlled by the range potentiometer RV. This control can therefore be calibrated in terms of radar range.

The second circuit we shall consider is illustrated in Fig 5. It is known as a *multiar* strobe pulse generator and is much used in RAF radar equipments. When the Miller timebase generator V_1 is cut off at its suppressor grid it is passing no anode current; V_1 anode is thus at h.t. +. The diode D_1 is then held *cut off* because its cathode is connected to h.t. + via T_1 secondary whilst its anode is returned to a lower voltage as decided by the range potentiometer RV. Under these conditions V_2 is conducting heavily, since its grid is returned to h.t. + via R_g , and V_{a2} is at a low value.

When the Miller rundown takes place, D_1 cathode voltage falls, and when the level set by RV is reached, D_1 conducts. V_2 grid is now connected via the conducting diode to the Miller

timebase and the Miller rundown causes V_2 grid voltage to fall. The resulting fall in current through V_2 and T_1 primary causes a voltage to be developed across T_1 secondary. The transformer is so connected that the cathode of D_1 and hence V_2 control grid are driven *negative*. This positive feedback produces a rapid cumulative action which quickly cuts off V_2 . As a result a *large positive-going step* is produced at V_2 anode, which rises to h.t.+. This positive-going step may be used as a strobe pulse step marker. Alternatively, the output may be differentiated to produce a large positive-going marker pip.

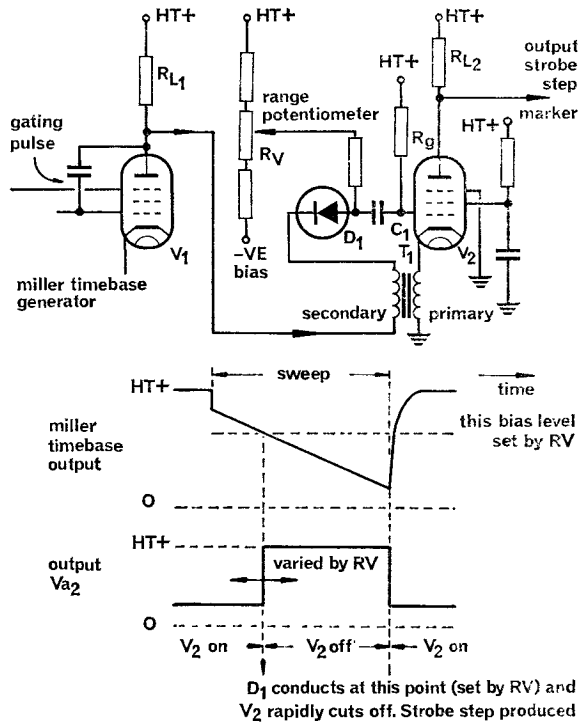


FIG 5. MULTIAR STROBE PULSE GENERATOR

The *position* of the strobe on the timebase trace depends upon the instant of time that V_2 is cut off and this, in turn, depends upon the setting of RV. Thus the strobe can be moved to any point on the timebase trace. As before, the range potentiometer RV is usually calibrated in terms of radar range.

Phase-shift Strobe Pulse Generator

The two strobe pulse generators considered in the preceding paragraphs depend for their operation on the sawtooth *timebase* voltage applied as an input. If this sawtooth waveform is non-linear in any way (which it is almost bound to be) the accuracy of ranging by means of the strobe pulse deteriorates. The *phase-shift* strobe pulse generator is a more accurate ranging system.

If we could arrange for a crystal-controlled oscillation to start every time the radar transmitter fired, the *absolute phase* of that oscillation at the time of the *received* echo would be an accurate measure of target range. Knowing the frequency of the oscillations we could count

the *number* of cycles and *fractions* of a cycle between the firing of the transmitter and the reception of the echo. Converting this into time, the range then follows. In practice it is not necessary to count *every* cycle. The sawtooth timebase waveform is normally linear enough to measure the time interval to the *nearest complete cycle*. All we then need to measure is the *phase*, within the limits of 0° – 360° , to give us the *fraction* of a cycle.

In this method of strobing, we need a phase-changer to provide a variable-phase signal. A resolver synchro is typical of the phase-changers used. In this application, sinusoidal voltages, equal in amplitude and frequency but differing in phase by 90° , are applied to the *stator* coils (Fig 6). This produces a magnetic field of constant amplitude, but the *phase* of the voltage induced in the *rotor* varies with its angular position. When the rotor is aligned with stator 1 the voltage induced in the rotor has the same phase as that in stator 1. Turning the rotor through 90° so that it is now aligned with stator 2 causes the voltage in the rotor to be of the same phase as that in stator 2, i.e. 90° *out of phase* with that produced in the first position.

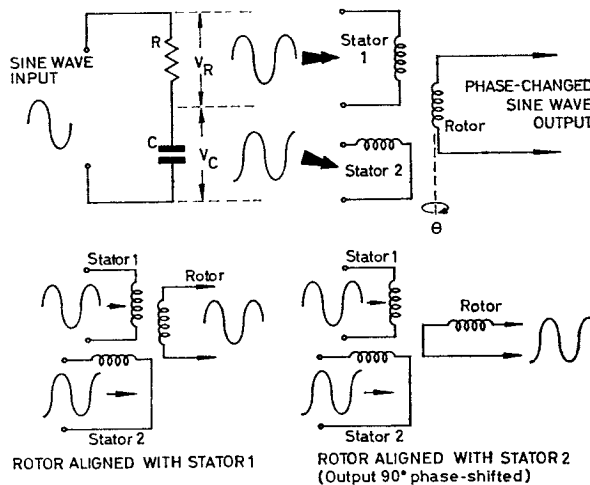


FIG 6. RESOLVER SYNCHRO AS A PHASE-SHIFTER

In this way it can be shown that rotating the rotor through 360° varies the phase of the induced voltage through 360° ; a 1° movement of the rotor produces a 1° phase shift.

Fig 7a illustrates the block diagram of a system using a phase-shift method of strobing. Fig 7b shows the associated waveforms. The trigger pulse from the master timing unit is applied to the gating valve V_1 , which produces a *wide* positive-going pulse. Its *pulse duration* is made *equal* to the *sweep time* of the range trace. In this example we have assumed a 0-10 miles range, so that the sweep time is $10 \times 10.75 = 107.5 \mu\text{s}$. The $107.5 \mu\text{s}$ gating pulse is applied to a crystal-controlled sine-wave oscillator and also to the timebase generator.

The sawtooth output is applied to the c.r.t. to produce the timebase trace. It is also applied to a multiar strobe pulse generator V_3 which, as we have seen, produces a narrow pulse during the period of the sweep. The *position* of this pulse in time relative to the start of the sweep is determined by the range potentiometer RV. The multiar output is used to trigger a pulse generator V_4 which produces a narrow pulse whose pulse duration is made equal to the period of *one cycle* of the sine-wave output of the oscillator. In this example this is assumed to be $10.75 \mu\text{s}$, equivalent to a radar range of one mile. This $10.75 \mu\text{s}$ pulse is applied as one input to the grid of a pentode coincidence gate V_7 .

The crystal-controlled oscillator V_5 oscillates for the period of the wide gating pulse, i.e. for $107.5 \mu\text{s}$. In this example the oscillator operates at a frequency of 93 kc/s so that the period

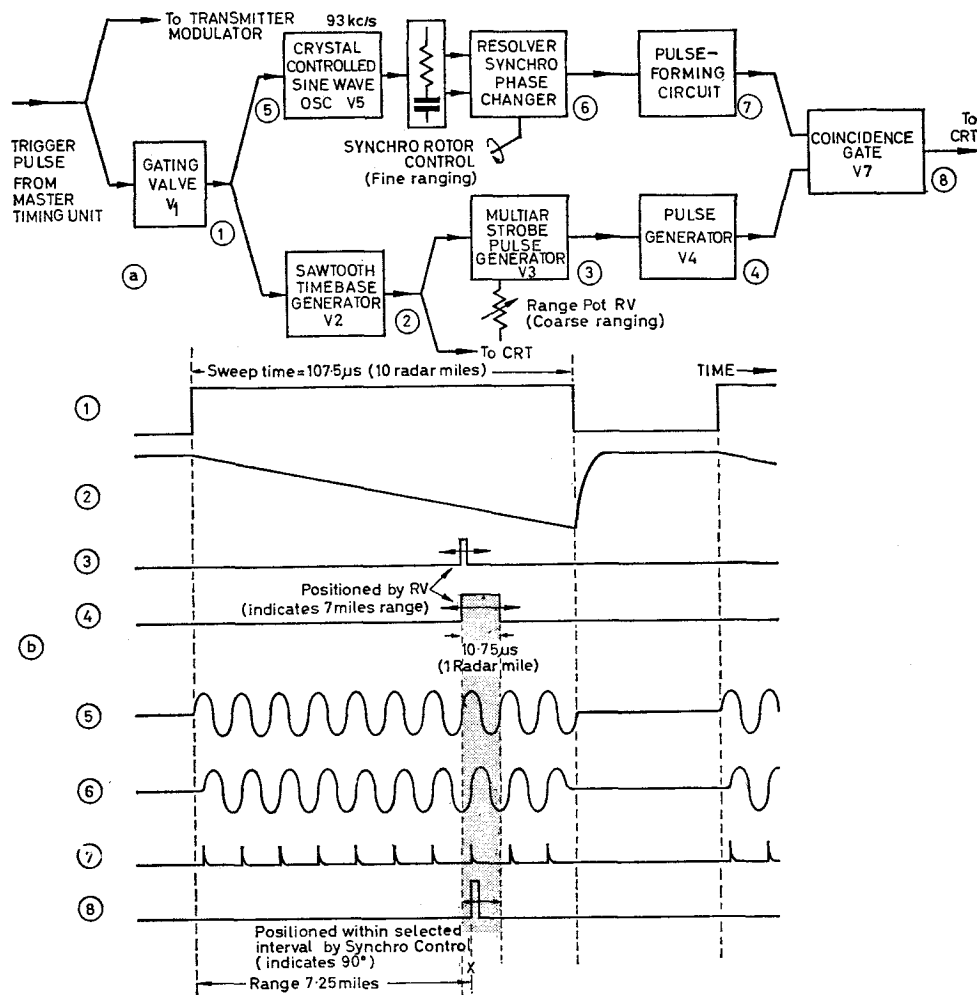


FIG 7. PHASE-SHIFT STROBE PULSE GENERATOR

of one cycle is $10.75 \mu\text{s}$, corresponding to a radar range of one mile. During the sweep time, *ten* cycles of sine wave are generated. The output from the crystal-controlled oscillator is applied to a network which produces two sine waves, 90° out of phase, for application to the stator windings of a resolver synchro. The phase of the output sine wave from the *rotor* depends upon the orientation of the rotor and this can be varied. The *phase-shifted* sine-wave output is applied to a pulse-forming circuit V₆ which produces one *very narrow* pulse at the instant each cycle of sine wave passes through zero in a positive-going direction. This train of pulses is applied as the second input to the suppressor grid of the pentode coincidence gate V₇.

The coincidence gate produces an output pulse only when its *two* inputs are *coincident*. This occurs at point X in Fig 7b. This final strobe pulse is applied to the c.r.t. as a strobe marker which can be moved to *any point* on the trace to mark the required target. This system ensures a very high degree of range accuracy. We have seen that the $10.75 \mu\text{s}$ pulse output from V₄ corresponds to a *range interval* of one mile. The time occurrence of this pulse within the sweep period depends upon the output from the multiar V₃ and this, in turn, is controlled by the range potentiometer RV. Thus *any given one-mile range interval* may be selected by

adjusting RV, and the range potentiometer will indicate this range as *so many miles*. The actual strobe marker pulse is then positioned very accurately *within* the selected one-mile range interval by varying the resolver synchro control to give *precise fractions* of a mile.

What we have in effect done is to 'count' the number of *complete* cycles of sine wave by means of RV to give the range to the nearest mile and then to measure the phase-shift corresponding to the *fraction* of a cycle left over to give very accurate or 'fine' ranging.

Auto-follow

In certain radar equipments it is necessary to generate a strobe pulse which *automatically follows* the movement of the selected target along the trace. This system is used, for example, in a modern AI equipment where the closing range of a target is measured automatically by a strobe marker pulse. When the strobe pulse indicates a certain pre-determined range it initiates the action of other circuits so that the fighter's armament is automatically brought into operation.

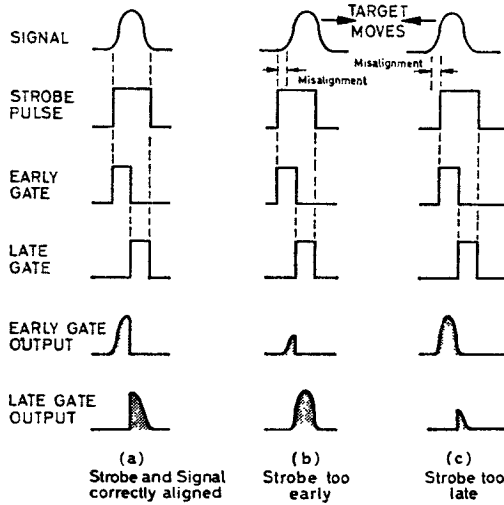


FIG 8. BASIC IDEA OF AUTO-FOLLOW

In one system, a strobe marker pulse is aligned with a target echo. The strobe pulse and the signal are then *coincident* in time. This marker pulse also operates two gates, one known as the *early gate* and the other as the *late gate*. The early gate opens coincident with the leading edge of the marker pulse, and as the early gate shuts the late gate opens. The pulse durations of the early and late gates are *equal*. The signal is routed via these gates and if the strobe marker pulse is correctly aligned with the target, the outputs from the early and late gates are *equal* (Fig 8a). If incorrectly aligned, as in Fig 8b and c, the outputs are *unequal*. Under this latter

condition a servomechanism system operates and moves the strobe marker pulse in such a direction as to make the outputs of the early and late gates again equal. In this way the strobe marker pulse is always *aligned* with the selected target echo. As the range of the target changes so does that of the strobe marker pulse. Change in range is thus automatically noted and appropriate circuits may be operated by the strobe pulse as required.

Strobe Markers for PPI Displays

In most radar equipments there is a short time interval between the end of one sweep of the timebase and the beginning of the next. This is normal to prevent ambiguities in range measurement (see p 37). This short time interval between successive traces is referred to as the 'inter-trace period' (Fig 9). In p.p.i. displays the inter-trace periods may be used for the production of strobe markers.

The simplest p.p.i. strobe marker is a bright spot which can be moved to any part of the display independently of the main

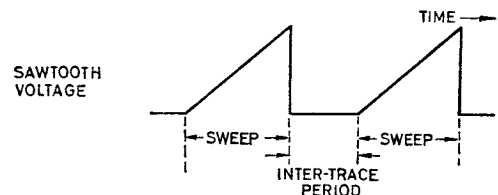


FIG 9. INTER-TRACE PERIOD

timebase trace position. Fig 10 illustrates the principle. During the sweep period the normal sine and cosine modulated sawtooth current waveforms are applied to the X and Y deflecting coils. A normal trace results, its radial position at that time depending upon the relative amplitudes of the X and Y sawtooth currents. During this time the blanking pulse to the c.r.t. grid ensures that the trace is just visible.

At the end of the sweep, the blanking pulse cuts off the c.r.t. beam at the grid so that flyback is not seen. For a very short period of time during each inter-scan period, pulses of current from the strobe pulse generator are applied to the X and Y deflecting coils. During this period the blanking pulse again returns the c.r.t. to normal conduction so that a bright marker spot is produced on the screen of the c.r.t. The *position* of this spot depends upon the relative amplitudes and polarities of the current pulses applied to the X and Y coils and this can be varied as required. The strobe spot marker can therefore be used to mark the position of any target on the display.

More complex marker systems may be used during the inter-scan period. It is possible to use several markers, different inter-scan periods being used for each one. Synthetic radar displays are often constructed in this way. We shall see more of this in Section 8.

Strobe Timebase

A strobe timebase enables an expanded version of a *small part* of the main timebase trace to be reproduced, sometimes on a separate c.r.t. and sometimes in place of the main timebase trace. This enables a target appearing on the main trace to be examined in much more detail.

For example, a target appearing on a 0-80 mile timebase will occupy only a very small part of the complete display. But if that part can be transferred in an *expanded* form to another trace, the target will appear in greater detail and may in fact show that more than one target is present (Fig 11).

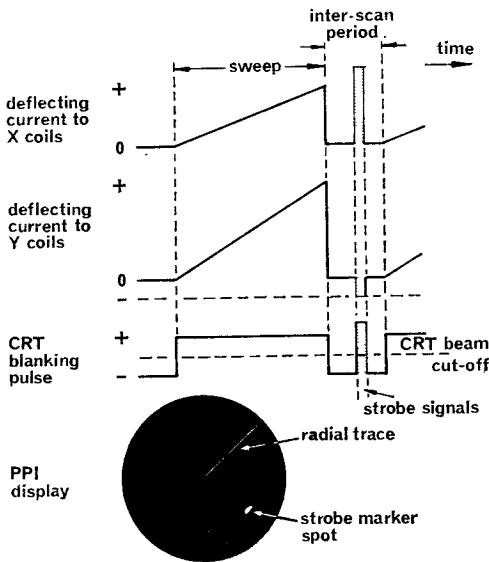


FIG 10. PRINCIPLE OF PPI STROBE SPOT MARKER PRODUCTION

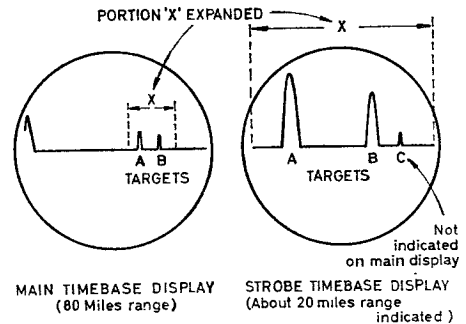


FIG 11. USE OF STROBE TIMEBASES

In some radar equipments the selected part of the main timebase trace is first marked by the strobe pulse. The main timebase generator is then switched off and the strobe timebase takes over to provide an expanded picture of the selected part of the trace.

In other equipments the main timebase trace is displayed continuously, and when a part of the main trace is to be examined in more detail that part is marked by a strobe pulse. The strobe pulse also triggers the strobe timebase generator which produces the expanded version of that part of the main trace under examination on a separate monitor c.r.t.

Fig 12 shows the block schematic arrangement of one possible system using separate main

and strobe displays. Stages 1 to 4 produce the strobe marker pulses and are as described earlier in this chapter. The strobe pulse, which is locked to the trailing edge of waveform 3, is applied to stage 5. The resulting square wave output is of fixed pulse duration but its *leading* edge is locked to the strobe pulse. Thus the pulse as a whole (referred to as a 'box' or 'pedestal' strobe pulse) can be moved to any point on the main timebase trace to mark the portion of trace which is to be expanded. The box pulse is also used as the gating pulse for the Miller strobe timebase generator. The output from this has the same *amplitude* as the main timebase output (i.e. the

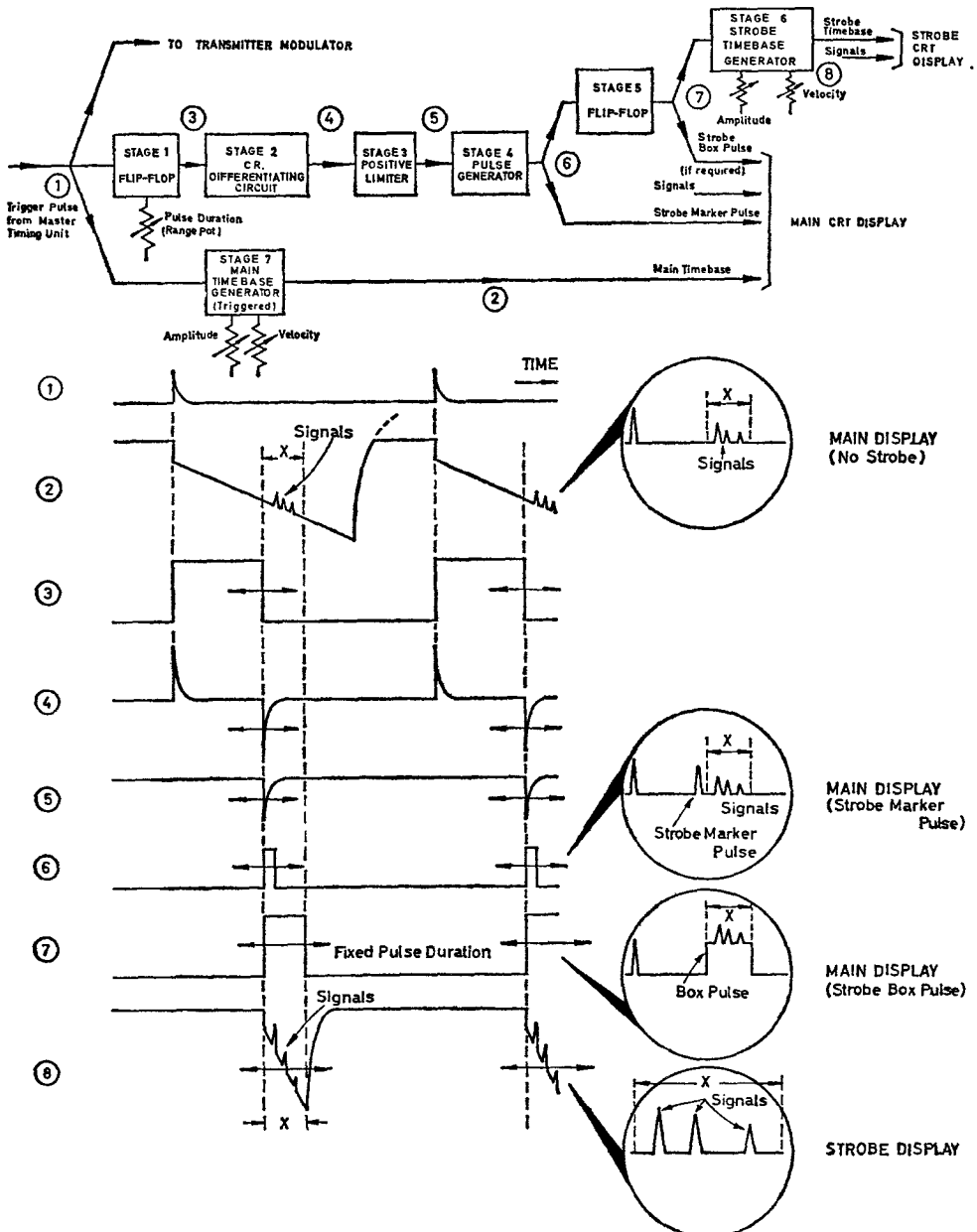


FIG 12. PRODUCTION OF STROBE TIMEBASE FOR TYPE 'A' DISPLAY

trace length is the same) but its *velocity* is much higher (i.e. the trace occurs in a much shorter period of time).

The part of the main timebase trace to be strobed is controlled by waveform 7, the leading edge of which is locked to the trailing edge of waveform 3. Thus although the strobe timebase in this example is of fixed duration the instant at which it operates is determined by the pulse duration control or range potentiometer in stage 1. In Fig 12 the portion of the main timebase trace being examined in more detail is denoted by 'X'.

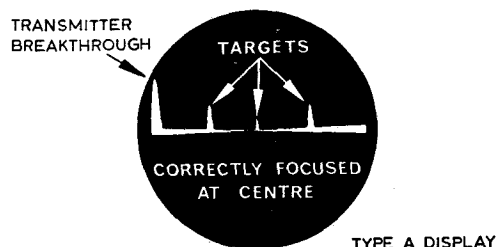
There is a great variety of strobe systems, most being peculiar to the actual equipment in which they operate. However, the information contained in this chapter covers most of the basic points we are liable to meet.

PARAPHASE AMPLIFIERS

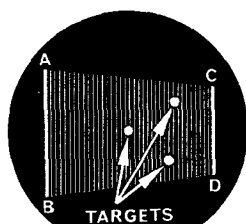
Introduction

All the timebase arrangements considered so far have assumed *single-ended* deflection in the c.r.t., i.e. one of the deflecting plates (or coils) is earthed and the deflecting voltages are applied to the other plate (or coil) in each pair.

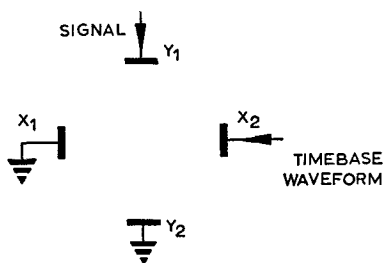
However, single-ended deflection can cause distortion, and is not normally good enough to be used in a radar indicator. Two common forms of distortion in electrostatic c.r.t.s are illustrated in Fig 1.



(a) DEFLECTION DEFOCUSING



(b) TRAPEZIUM DISTORTION



(c) SINGLE-ENDED DEFLECTION, TYPE A DISPLAY

Causes of Distortion

In an electrostatic c.r.t., the final anode, which is part of the focusing system, is normally earthed. The *mean* voltage of the deflecting system should also be zero. However, when we apply the timebase voltage to only *one* of the deflecting plates, the

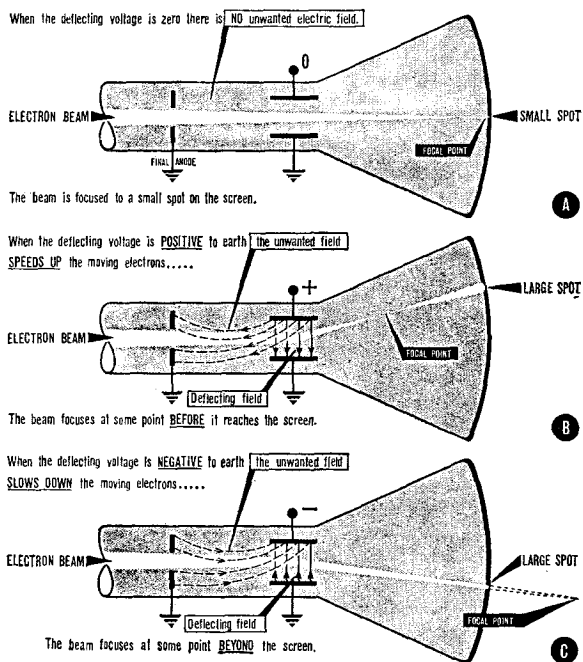


FIG 1. DISTORTION IN ELECTROSTATIC CRTS, SINGLE-ENDED DEFLECTION

FIG 2. EFFECT OF SINGLE-ENDED DEFLECTION, DEFLECTION DEFOCUSING

mean voltage of the deflecting system relative to earth *changes*. The required deflecting field is set up between the plates. But we also get an *unwanted* field between the plate to which the voltage is applied and the final anode, as shown in Fig 2b and 2c.

Because the unwanted field is mainly parallel to the direction of the beam, the change in the mean voltage of the deflecting system either *speeds up* or *slows down* the moving electrons.

This moves the *focal point* of the beam at the screen so that the beam is no longer focused to give a small spot. Whatever the polarity of the deflecting voltage, the effect on the screen is that the spot grows larger as it moves from the *centre* of the screen. Thus when we apply a timebase waveform to the X deflecting plates we get the effect on the screen which we saw in Fig 1a. Because the spot defocuses when it is deflected from the centre, this form of distortion is known as *deflection defocusing*.

Trapezium distortion occurs in a similar way. If X_1 is earthed and X_2 is made *positive* with respect to X_1 , the beam is deflected to the right. At the same time however the *mean* voltage of the deflecting system has increased and this *increases* the speed of the electrons in the beam. Thus the deflection produced by the Y plates for a given deflection voltage *decreases* and a short trace CD results (Fig 1b). Conversely, if X_2 is made *negative* to X_1 , the decrease in electron velocity causes an *increase* in the Y plate deflection sensitivity; a long trace AB results (Fig 1b). The deflection sensitivity of the X plates is affected in the same way by deflecting voltages applied to the Y plates.

Balanced Deflection

If, instead of using single-ended deflection, we use *balanced* or *symmetrical* deflection, both forms of distortion discussed above are reduced. For example, instead of applying $+10V$ to one plate of a pair with the other earthed, let us apply $+5V$ to one and $-5V$ to the other. The voltage *between* the plates remains the same so the amount of deflection is the same in each case. However in the second case there is *no change in the mean voltage* of the deflecting system. Thus the speed of the electrons in the beam remains constant and distortion is reduced. Fig 3 shows that there is now no unbalancing field between the deflection plates and the final anode.

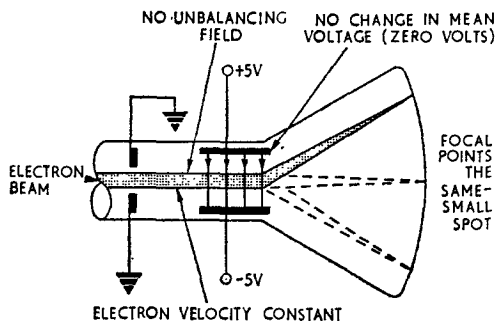


FIG 3. EFFECT OF BALANCED DEFLECTION

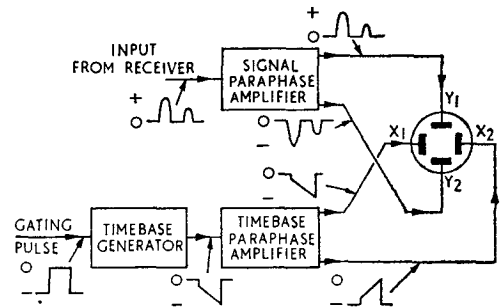


FIG 4. USE OF PARAPHASE AMPLIFIERS IN A TYPE 'A' DISPLAY

This method of using two *equal* and *opposite* deflecting voltages for each pair of plates is used in most radar indicators. For balanced deflection we need two deflecting waveforms which are 180° *out of phase* or *paraphase*. For sine waves, paraphase voltages could be obtained by using a transformer with a centre-tapped secondary. For square wave and sawtooth inputs, however, the distortion resulting from using transformers would be excessive. For such inputs a circuit known as a *paraphase amplifier* is used instead.

In a radar indicator using a type A display we need balanced deflection on *both* pairs of deflecting plates (X and Y). We therefore use *two* paraphase amplifiers—one for the timebase input and one for the signal input (Fig 4).

For minimum distortion the timebase and signal waveforms should be balanced with reference to earth. This can be obtained with the simple phase-splitting system shown in Fig 5a. The waveforms show that one output is *in phase* with the input and the other 180° *out*

of phase. Notice that since the input and output 1 are in phase, the input may be used directly to give this output, as in Fig 5b. In this arrangement the system merely provides a 180° phase reversal *without amplification*.

Single Valve Paraphase Amplifier

Fig 6 shows the circuit of an amplifier which provides unity gain and the normal 180° phase reversal. Thus it corresponds to the arrangement of Fig 5b.

The input to the valve grid is from the potential divider R_1R_2 and has a value:

$$V_g = V_{in} \frac{R_2}{R_1 + R_2}.$$

The ratio $\frac{R_2}{R_1 + R_2}$ is chosen to be $\frac{1}{A}$, where A equals the gain of the stage. Thus if the stage has a gain of 40, R_2 could have a value of 10 kΩ and R_1 a value of 390 kΩ. $\frac{R_2}{R_1 + R_2}$ would then have the correct value of $\frac{1}{40}$. Amplification by the stage would produce an output of $40 \times \frac{V_{in}}{40} = V_{in}$. Thus the output is equal in amplitude to the input with a phase reversal of 180°.

Normally R_k is left un-decoupled. This provides *current negative feedback* to reduce the effects of variations in the supply voltage and in the circuit.

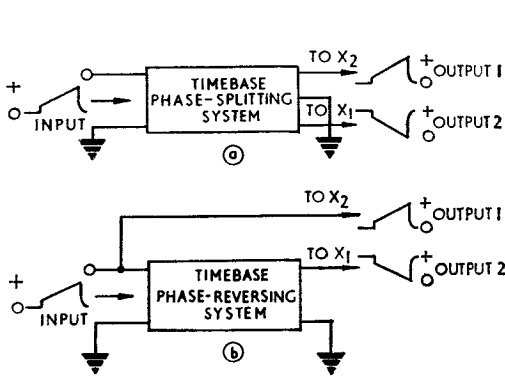


FIG 5. TYPICAL PARAPHASE SYSTEMS

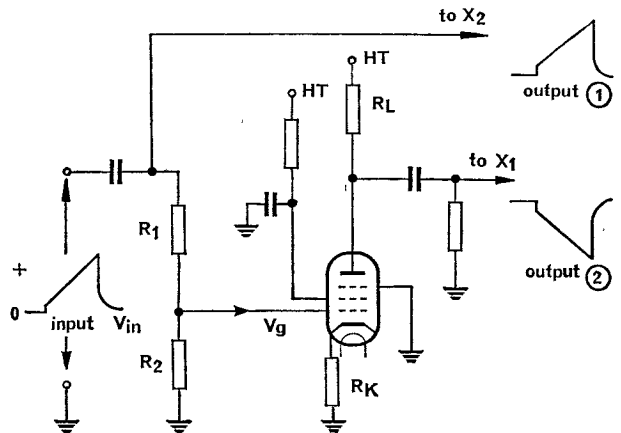


FIG 6. SINGLE VALVE PARAPHASE AMPLIFIER

Two Valve Paraphase Amplifier

Where the output from the timebase generator is of insufficient amplitude to give the required trace length, it is necessary for the paraphase amplifier to provide *gain* in addition to phase-splitting. Two stages are then necessary. The easiest arrangement is to insert a high gain RC amplifier before the single valve paraphase amplifier of Fig 6. This is shown in Fig 7.

The potential divider R_1R_2 must be such that the output from V_1 , which serves as the input to V_2 , is reduced by A_2 times, where A_2 is the gain of V_2 . Thus the action is the same as that of Fig 6 except that *both* outputs have been amplified A_1 times, where A_1 is the gain of V_1 .

Phase-splitter

This is basically an amplifier in which the load is split into *two equal halves*, one in the collector and the other in the emitter (Fig 8). The circuit thus acts as a combination of amplifier and emitter-follower and corresponds to the system of Fig 5a.

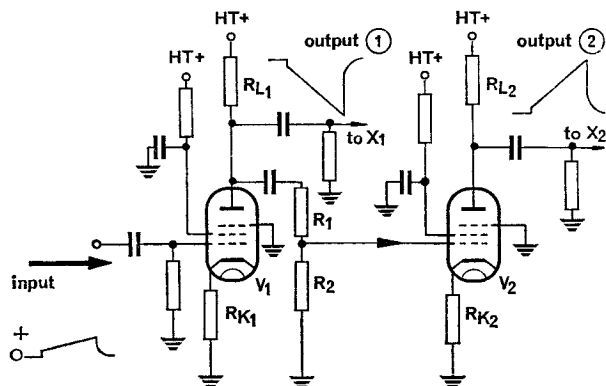


FIG 7. TWO VALVE PARAPHASE AMPLIFIER

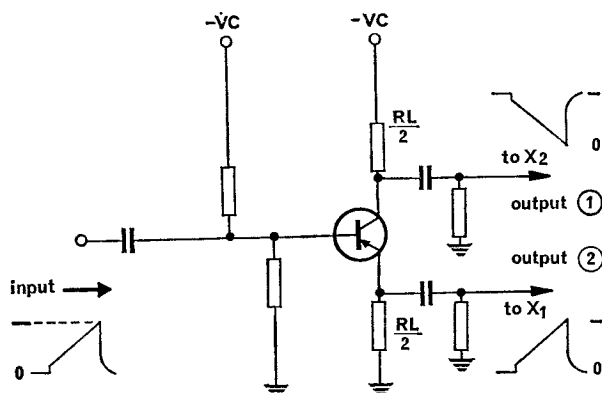


FIG 8. BASIC PHASE-SPLITTER

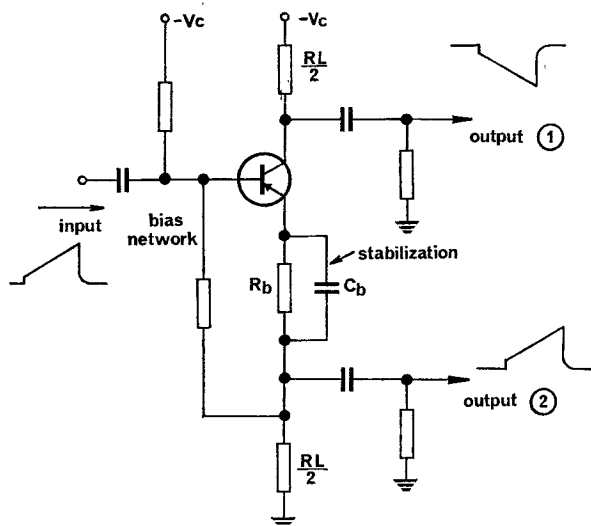


FIG 9. BIAS ARRANGEMENTS IN PHASE-SPLITTER

A negative-going input produces a positive-going output at the collector and a negative-going output at the emitter. Since the load in each output lead is the same, the two outputs are equal in amplitude. Because of the emitter-follower action neither output can be greater than the input (see AP 3302, Part 1B, p 310).

The circuit has the normal emitter-follower advantages of good linearity and stability. Miller effect is negligible and in a valve circuit a triode may be used even at high frequencies since the gain is less than unity.

A practical value for $\frac{R_L}{2}$ would be about 10 k Ω so that even a small current would develop an appreciable voltage drop. The result might be too much bias between emitter and base. This may be overcome as shown in the circuit of Fig 9. In this circuit the base resistor is returned to the junction of R_b and $\frac{R_L}{2}$ in the emitter. This ensures that the d.c. voltage drop across the emitter load $\frac{R_L}{2}$ is *not applied as bias* to the base.

Long-tailed Pair

The long-tailed pair is mentioned briefly in Part 1B of these notes (p 310). It is a two-stage paraphase amplifier in which the first stage acts as a phase-splitter and the second stage as a grounded-grid amplifier (Fig 10a). The basic action of the phase-splitter is as previously described. The output from the anode is in *anti-phase* with the input whilst the output from the cathode is *in phase* with the input. The cathode output is applied as input to the grounded-grid stage. There is *no phase inversion* in a grounded-grid circuit. Thus a positive-going input applied to the *cathode* of V_2 from R_K appears as a positive-going output at V_2 anode.

Combining these two circuits gives the long-tailed pair shown in Fig 10b. The waveforms show that a positive-going input at V_1 grid produces an amplified

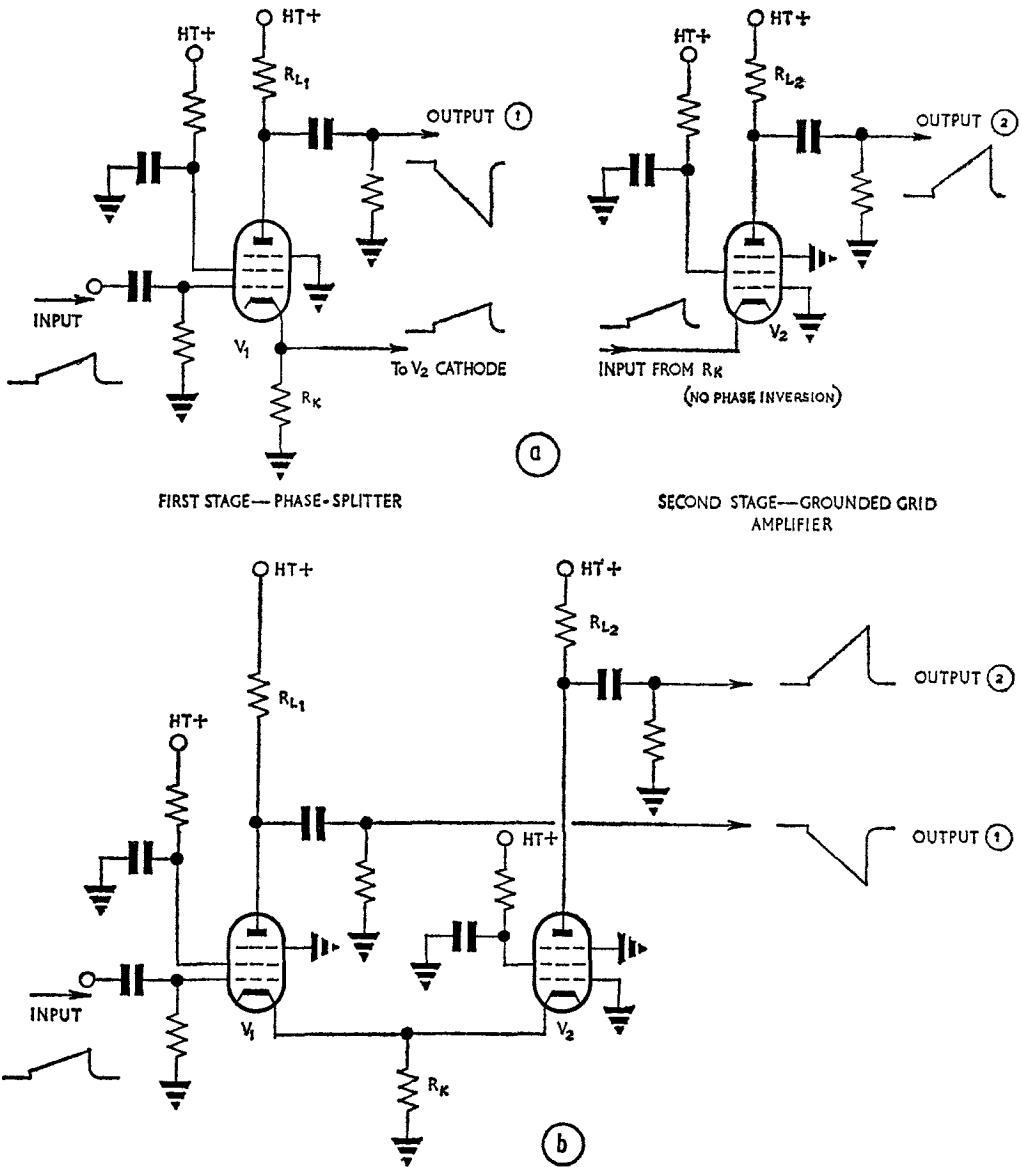


FIG 10. LONG-TAILED PAIR

negative-going output from V_1 anode and an amplified *positive-going* output from V_2 anode.

Since the output voltages are in *anti-phase*, the two valve currents are in anti-phase through R_K . Thus if we had exactly equal anode loads and equal valve currents there would be *no resultant change* in the current through R_K . Thus the voltage across R_K would remain constant and there would be *no input* to V_2 . Hence this circuit can operate only if there is a *difference* between the two outputs. Provided R_K has a high value, the required current change through it need only be *very small* to give an adequate input to V_2 . Thus a very small difference between the two outputs would produce the required results. In addition, with such a small cathode

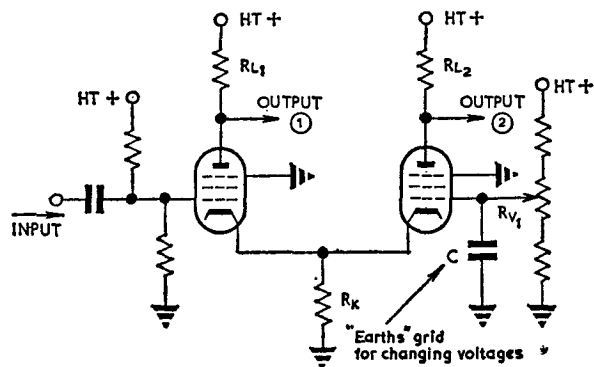


FIG 11. BIAS ARRANGEMENTS IN LONG-TAILED PAIR

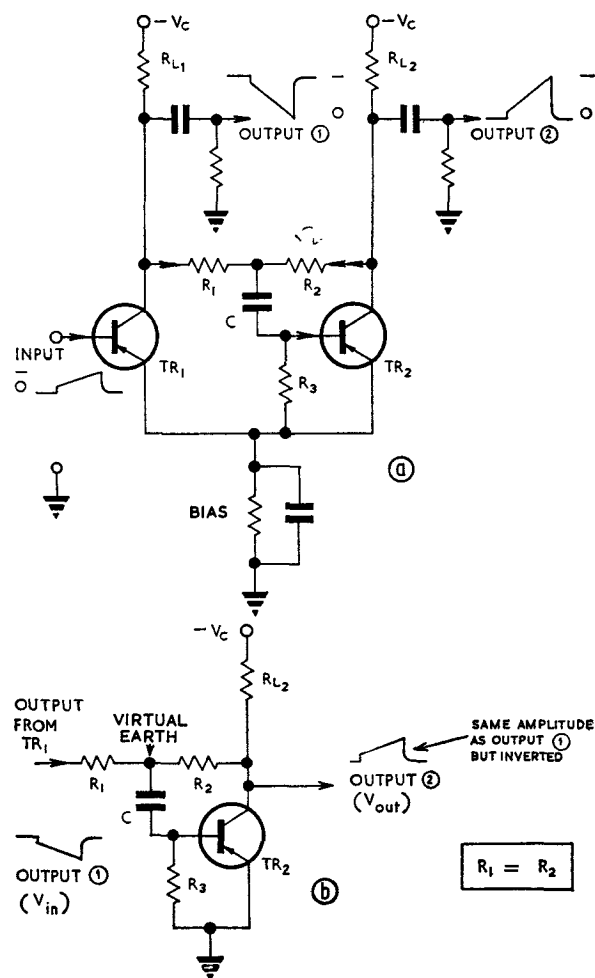


FIG 12. FLOATING PARAPHASE AMPLIFIER

input to V_2 , the g_m of this valve needs to be high. Given *high* values of R_K and g_m the anti-phase outputs will have *almost equal* amplitudes.

R_K normally has a value of several kilohms. In the basic circuit of Fig 10b, the voltage across R_K is also the *bias* voltage on the two valves. Since R_K has a high value some modification to the basic circuit is needed to prevent an excessive bias. To reduce the effective bias, R_K may be returned to a negative bias point. This has the effect of making the grids *more positive* with respect to the common cathode. Alternatively, a positive voltage may be applied to the grids from voltage divider networks connected between h.t.+ and earth, as shown in Fig 11. If we vary the bias on one of the grids (e.g. by RV_1 in Fig 11), and use d.c. coupling between each anode and the appropriate deflector plate, we have a *balanced shift control*. Thus if V_{g_2} is made more positive, V_{a_2} falls; V_{R_K} will rise and so will V_{a_1} . Thus we have anti-phase shift voltages at each anode.

Floating Paraphase Amplifier

The basic circuit of a floating paraphase amplifier is shown in Fig 12a. It can be split into two sections: the amplifier TR_1 and the stage TR_2 (with the network R_1 , R_2 , C and R_3). The stage TR_2 , shown separately in Fig 12b, is a *see-saw amplifier* (see p 575 of AP 3302 Part 1B). The see-saw amplifier has a high gain and uses a large amount of shunt voltage negative feedback via R_2 . If the feedback resistor R_2 and the input resistor R_1 are *equal*, the output voltage is *equal in magnitude* but of *opposite polarity* to the voltage applied to R_1 . The see-saw amplifier therefore acts as an *inverter*. Note that the junction of R_1R_2 , connected through C to the base of TR_2 , is at 'virtual earth' because of the symmetrical input and output voltages.

It is easy to show that when R_1 equals R_2 , V_{out} equals V_{in} . If the base

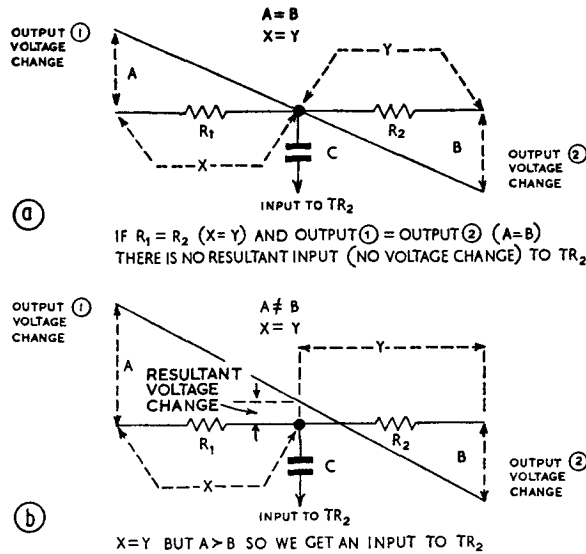


FIG 13. NEED FOR SLIGHT CIRCUIT UNBALANCE

is at virtual earth TR_2 draws no base current and the current through R_1 must all flow through R_2 also. Thus V_{R_1} equals V_{R_2} . But V_{R_1} is the input voltage V_{in} and V_{R_2} is the output voltage V_{out} . Therefore, if R_1 equals R_2 , V_{in} equals V_{out} with the usual 180° phase change.

Let us now combine the two parts of the circuit. A *negative-going* input to TR_1 base is amplified and inverted to give a *positive-going* output from TR_1 collector. Output 1 also acts as the input to TR_2 via R_1 . As we have seen, the see-saw action of TR_2 then provides an equal amplitude *negative-going* output from TR_2 collector.

So far we have assumed that R_1 equals R_2 and output 1 equals output 2. If this were true there would in fact be *no input* to TR_2 and the stage would cease to operate (Fig 13a). For the

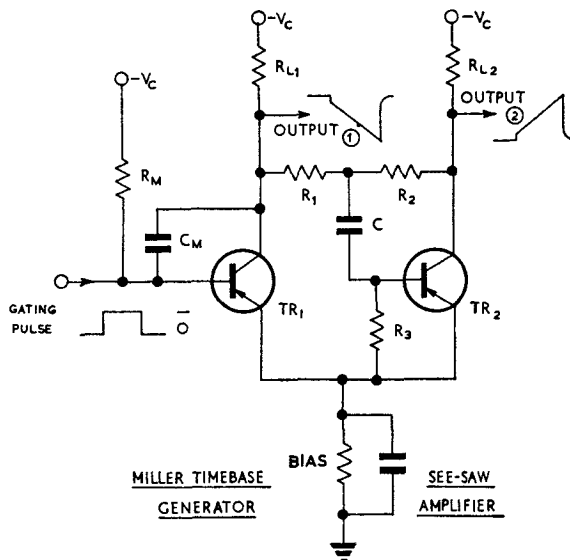


FIG 14. TYPICAL RADAR TIMEBASE ARRANGEMENT

circuit to work there must be a *slight unbalance* between the two halves of the circuit, just as for the long-tailed pair. In practice R_1 equals R_2 and the circuit settles down to the condition where output 2 is slightly *smaller* (or *larger*) than output 1 (with a 180° phase change).

This ensures a small input to TR_2 base (Fig 13b). If TR_2 is a high gain amplifier the unbalance need only be very small.

Many radar equipments use the floating paraphase amplifier. Negative feedback is applied only to TR_2 . Because of this it is usual to use the Miller timebase generator itself as the stage TR_1 . This ensures that no distortion is introduced by a separate amplifier. The arrangement is as shown in Fig 14.

Paraphase Amplifiers for Magnetic CRT

The paraphase amplifiers just considered are *voltage* amplifiers whose outputs may be used to give balanced deflection in electrostatic c.r.t.s. For magnetic c.r.t.s the paraphase amplifier is followed by a push-pull power amplifier which produces the required deflection currents. Equality of the output voltages is then less important because the power amplifiers can usually be adjusted to provide equal deflection currents.

SECTION 3
UHF RADAR

Chapter		Page
1	UHF Radar	227
2	UHF Radar Aerial Systems	237

CHAPTER 1

UHF RADAR

Introduction

We have seen in Section 1 of this book (see p 16) that radar frequencies in use extend from about 200 Mc/s to about 100,000 Mc/s. Table 1 on p 33 shows that these frequencies may be split up into *bands*.

The band around 200 Mc/s is the *P band*. The corresponding wavelength is 1.5m so that radars operating in this band are often referred to as *metric* radars. Note also that this band falls within the *v.h.f. range* of frequencies (30 to 300 Mc/s). Although metric radars operating at frequencies as low as 30Mc/s were common in the early days of radar, the applications at v.h.f. are now much less. The main reason for this is that large aerial structures are required at these frequencies to provide a narrow beamwidth. However, where narrow beamwidth is not a requirement, radars operating around 200 Mc/s can be used. The early warning search radar type 7 considered in Section 1 (p 44) is a metric radar. Some of the secondary radar navigational aids also use this frequency band (e.g. Rebecca).

Wavelengths corresponding to frequencies higher than 1,000 Mc/s are referred to as *microwaves*. Two of the most commonly used bands in the microwave region are the *S band* (3,000 Mc/s, 10 cm) and the *X band* (10,000 Mc/s, 3 cm). Radars operating in these bands are usually referred to as *centimetric* radars. The S band is used mainly for medium-range surveillance, whilst the X band is ideally suitable for short-range precision approach radars.

For even shorter range surveillance (e.g. airfield surface movement) the *Q band* (40,000 Mc/s, 7.5 mm) may be used. Radars operating in this band are usually referred to as *millimetric* radars.

Lying between the metric (v.h.f.) and the centimetric (microwave) bands we have the frequency spectrum of 300-1,000 Mc/s. This lies within the *u.h.f. band* (300-3,000 Mc/s) and includes the lower part of the *L band*. The corresponding wavelength is about 50 cm. Until about 1950 this band was virtually ignored for radar purposes for reasons which we shall see shortly. However, since then, a large amount of radar development has been carried out in the 50 cm u.h.f. region, especially for long-range radars. The result is that many high-power, long-range radars are now operating at these lower u.h.f. radar frequencies.

This section discusses u.h.f. radar. Chapter 1 deals with its advantages and limitations, and also the components used at u.h.f. Chapter 2 deals with u.h.f. aerial systems. Centimetric radars, operating at frequencies above 1,000 Mc/s, are discussed in Section 4.

Advantages of UHF Radar

1. **High transmitter power.** Higher transmitter powers are usually easier to achieve at the lower radar frequencies than at the higher frequencies. This follows from the fact that practically all losses (skin effect, capacitive and inductive losses, etc.) increase with increase in frequency. The *range* of a pulsed radar depends, among other things, upon the peak power during the pulse and this, in turn, is related to the *average* transmitter power. Thus, for long ranges we require a transmitter with *high* average power and, as we have seen, this can be obtained more easily at the lower radar frequencies. In addition, *c.w. radar* is now used for many applications and this also requires the use of transmitters with high average power.

2. **Good MTI performance.** We shall see later that moving target indication (m.t.i.) radars must operate at low frequencies, or with a high p.r.f., or both, to give satisfactory performance. MTI radars normally produce pulses of relatively long duration (e.g. 10 μ s), and to obtain a

reasonable peak power output during this time, transmitters with *high average power* are needed. The lower radar frequencies can provide this more easily. Many m.t.i. radars operate at u.h.f. in the region of 500 Mc/s.

3. Weather effects. We stated in Section 1 (p 30) that for radar detection of an object the wavelength in use must be comparable in size with the target. Thus for the detection of *small* objects the radar wavelength must also be small, i.e. the frequency must be *high*. At the higher radar frequencies the wavelength becomes comparable in size with rain and cloud particles and these particles then produce a radar echo signal. Thus we get considerable *precipitation clutter* at the higher radar frequencies. At frequencies higher than about 3,000 Mc/s the clutter due to weather effects may be so bad as to completely mask the desired target signal. Although precipitation clutter may be reduced by using circular polarization, the most positive way of eliminating unwanted weather effects is to operate at the *lower* radar frequencies. At frequencies up to 1,000 Mc/s weather clutter is negligible.

4. Atmospheric absorption. Energy is extracted from a radar signal in its passage through the atmosphere. The *amount* of energy extracted and absorbed by the gas molecules in the atmosphere depends upon *frequency* and, in general, increases with increase in frequency (i.e. with *decrease* in wavelength). If we drew a graph of atmospheric attenuation against frequency we would notice that, although the graph is rising with increase in frequency, it is *not a linear rise*. The graph contains peaks and troughs due to the 'resonance' effects of the gas molecules. Thus, for example, the attenuation is actually *less* at 100,000 Mc/s (0.3 cm) than at 60,000 Mc/s (0.5 cm). However, in general, there is very great attenuation of signal strength at wavelengths shorter than 3 cm (10,000 Mc/s) and these wavelengths can be used only for *short-range* working. At frequencies below 1,000 Mc/s atmospheric attenuation is negligible.

5. Long range. We have already seen that increased range can be expected at the lower radar frequencies because of the higher average transmitter power available at these lower frequencies,

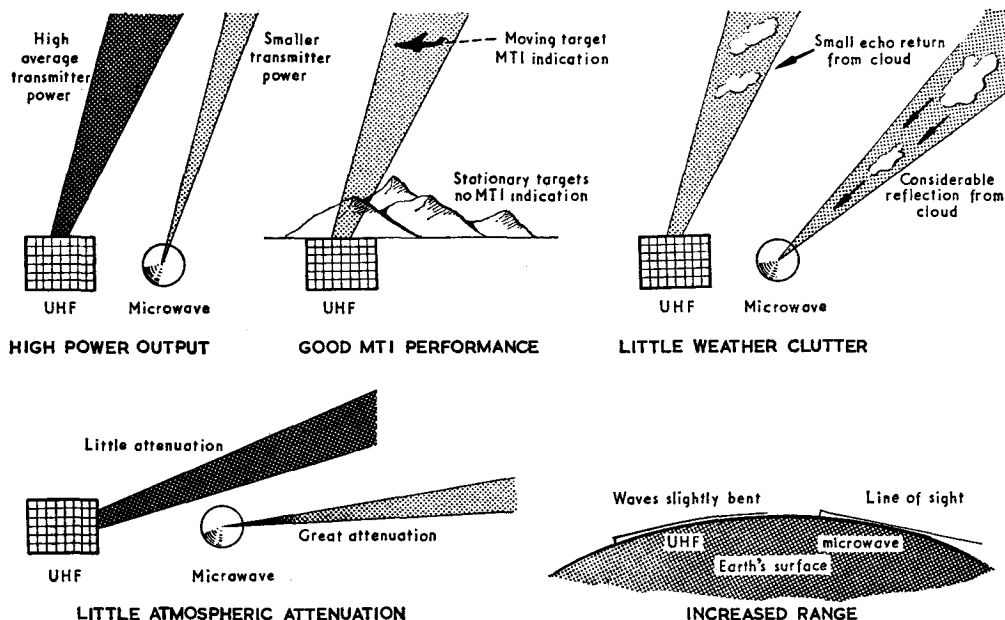


FIG 1. ADVANTAGES OF UHF RADAR

and because of the reductions in weather clutter and atmospheric attenuation. There is however another factor. At the lower radar frequencies the radar waves are 'bent' slightly in the atmosphere so that they tend to follow the earth's curvature. Thus if low-altitude radar coverage is required *beyond* the geometrical horizon, the frequency should be as low as possible. Extremely long range radars are essential for early warning of ballistic missiles. For this and other reasons, ballistic missile early warning systems (BMEWS) operate in the u.h.f. band around 500 Mc/s.

Fig 1 summarizes the advantages of u.h.f. radar outlined on previous pages.

Disadvantages of UHF Radar

1. **Aerial size and beamwidth.** We have already seen that to produce a beam of radiation some form of aerial array is required. For a narrow beamwidth an aerial array *large in terms of wavelength* is needed. Thus, *very large* aerial systems are required at the lower radar frequencies to provide a narrow beamwidth. At 70,000 Mc/s a 1° beamwidth can be obtained with a parabolic reflector 1 foot in diameter. At 300 Mc/s a diameter of about 200 feet would be required for this beamwidth. Although both these aerials have the same gain, the signal input to the receiver is greater for the low-frequency aerial. This is because the power delivered by a receiving aerial to its matched load depends upon the product ($\text{gain} \times \lambda^2$), i.e. proportional to *aerial aperture*. Thus, if we are thinking only in terms of the receiver signal input we can afford to make the aerial *smaller* than 200 ft diameter—although of course the *beamwidth* will then increase and accuracy of angular measurement decrease.

2. **Angular measurement accuracy.** With aerials of a practicable size the beamwidth at the lower radar frequencies is greater than that at the higher frequencies. Angular measurement accuracy and angular discrimination are therefore poorer at the lower frequencies.

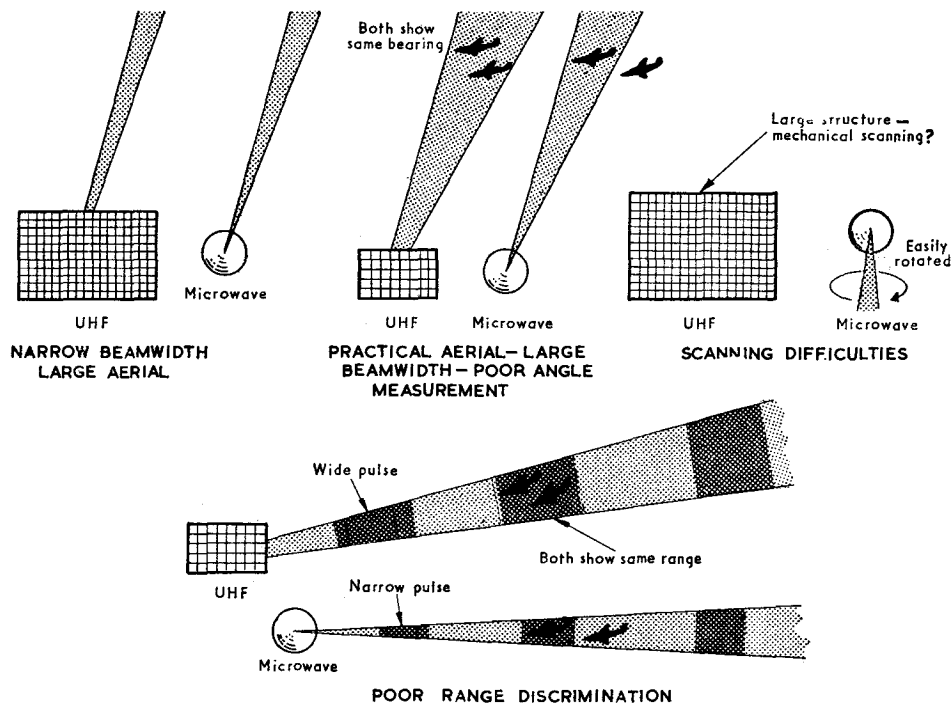


FIG 2. DISADVANTAGES OF UHF RADAR

3. **Scanning problems.** With the large aerial arrays that are essential at the lower radar frequencies, mechanical scanning methods introduce problems. Fortunately, as we shall see, it is possible to 'steer' a beam by *electronic* means without having to move large mechanical structures. This is the method normally used in large ground radars operating around 500 Mc/s.

4. **Range accuracy and resolution.** For satisfactory reflection of energy from a target, each radiated pulse must contain an adequate number of r.f. cycles. At low frequencies a pulse of *long duration* is needed to give this and, as we noted in Section 1 (p 39), target discrimination in range then becomes poor.

Fig 2 summarizes the disadvantages of u.h.f. radar as outlined above.

Choice of Radar Frequency

The frequency chosen for a particular radar depends upon the job it has to do. Some factors are more favourable at the lower frequencies (e.g. freedom from weather effects, little atmospheric attenuation, longer ranges); others are more readily obtained at the higher frequencies (e.g. narrow beamwidth, good angular accuracy, good range resolution). Choice of frequency is therefore a *compromise*.

For the reasons given earlier, a high-power u.h.f. radar will provide extremely long ranges and give good m.t.i. performance. It cannot however produce good angular measurement accuracy or good range resolution. Thus the main role of the high-power u.h.f. radar is as an *early-warning* system, where accuracy is of secondary importance.

A 10 cm (3,000 Mc/s) S band radar suffers more than the u.h.f. radar from the effects of weather and from atmospheric attenuation. Its range is therefore less. However, because of the narrower beamwidth and shorter pulse durations, it can provide good accuracy and resolution. Radars in this band are therefore suitable for medium-range surveillance and height-finding (up to about 200 miles), with similar 'search' roles for airborne equipment.

A 3 cm (10,000 Mc/s) X band radar suffers severely from weather effects and from atmospheric attenuation. But its accuracy is very high. Thus it is ideal for precision approach radars, AI, and similar systems where a high degree of accuracy over only short ranges (e.g. 30 miles) is required.

A 7.5 mm (40,000 Mc/s) Q band radar is suitable only for very short range work (within about 5 miles) because of very heavy atmospheric attenuation. The extremely narrow beamwidth which can be produced at these wavelengths means that accuracy and resolution are very high. In fact, television-like pictures can be produced at short ranges. One of its main uses is the surveillance from the control tower of *aircraft movement on the ground*.

Tuned Circuits at UHF

We considered u.h.f. tuned circuits in Part 1B of these notes (p 488). There we stated that at u.h.f. the normal parallel LC tuned circuit becomes very 'lossy' and of very small physical size. Thus other forms of tuned circuit become necessary at frequencies higher than about 200 Mc/s.

At frequencies between about 200 and 400 Mc/s, *lecher lines* may be used in place of a normal LC tuned circuit (Fig 3a). These are lengths of open wire feeder so adjusted that they are resonant at the required frequency.

Coaxial, or *concentric*, *line resonators* can be used in much the same way as lecher lines (see Fig 3b). Because of their construction, radiation and other r.f. losses of coaxial line resonators are less than those of open wire lecher lines and they can be used at frequencies of the order of 1,000 Mc/s.

The waveguide type of *cavity resonator* may also be used at u.h.f. (see Fig 3c). The cavity resonator was mentioned in p 491 of Part 1B and it is considered in more detail in Section 4 of

this book. Although the cavity resonator is usually considered as a 'microwave' device it also finds application at frequencies as low as 300 Mc/s. In fact, the majority of tuned circuits in u.h.f. radar are of the cavity resonator type.

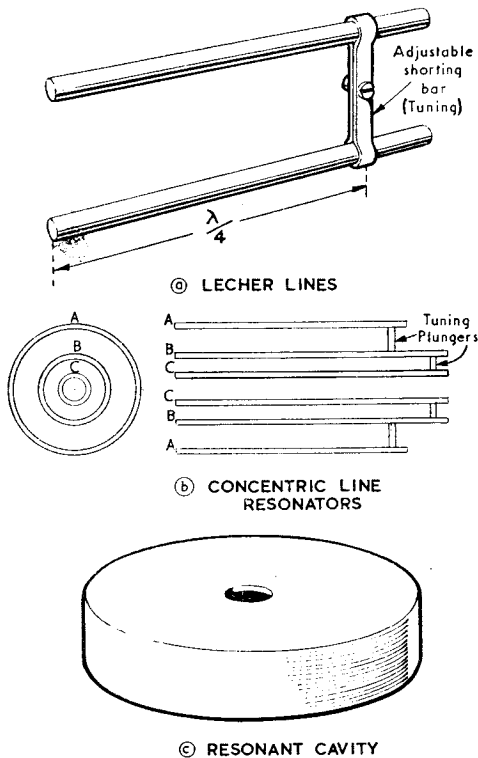


FIG 3. TUNED CIRCUITS AT UHF

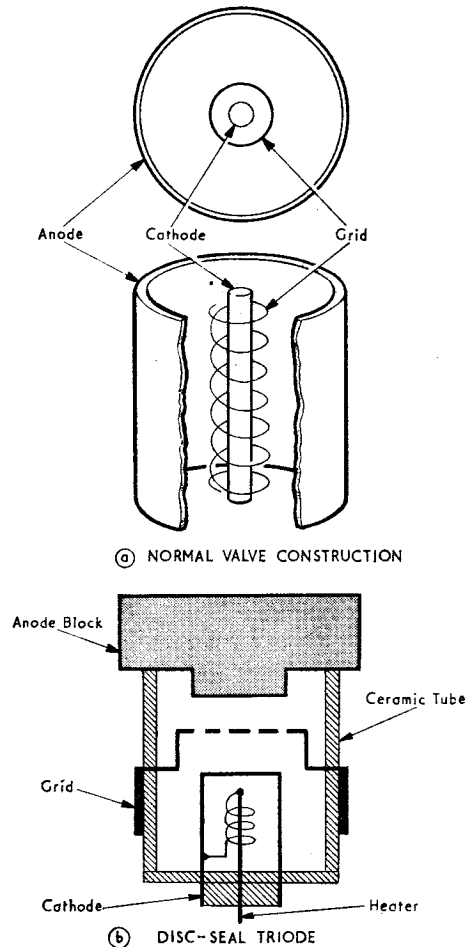


FIG 4. CONCENTRIC AND PLANAR FORM OF TRIODE CONSTRUCTION

Grid-controlled Valves at UHF

Valves for use at frequencies up to v.h.f. are constructed as shown in Fig 4a. In this type of construction the cathode is the central electrode and the grids and anode are in the form of concentric cylinders around the cathode. Because of the inductance of electrode connecting leads, inter-electrode capacitances, transit time effects and so on, this type of construction is unsuitable for u.h.f. and higher frequencies.

Grid-controlled valves for these frequencies must have flat, or *planar*, electrodes. The connections to the electrodes are taken straight out of the side of the valve through metal-to-glass, or metal-to-ceramic, disc seals. The *disc-seal triode* considered in p 488 of Part 1B, and illustrated in Fig 4b, is typical of this form of planar construction.

In the planar triode, the metal discs to which the electrodes are connected, form ideal terminations for coaxial line resonators. Radiation and lead inductance losses are therefore greatly reduced. In addition, the planar structure permits the use of small electrode areas which reduces inter-electrode capacitances. The electrodes can also be spaced very close together to reduce transit time effects.

A typical u.h.f. amplifier using a disc-seal triode connected to a concentric line resonator is illustrated in Fig 5a. The equivalent circuit, shown in Fig 5b, is seen to be a conventional grounded-grid r.f. amplifier. This arrangement can be converted into an oscillator by connecting

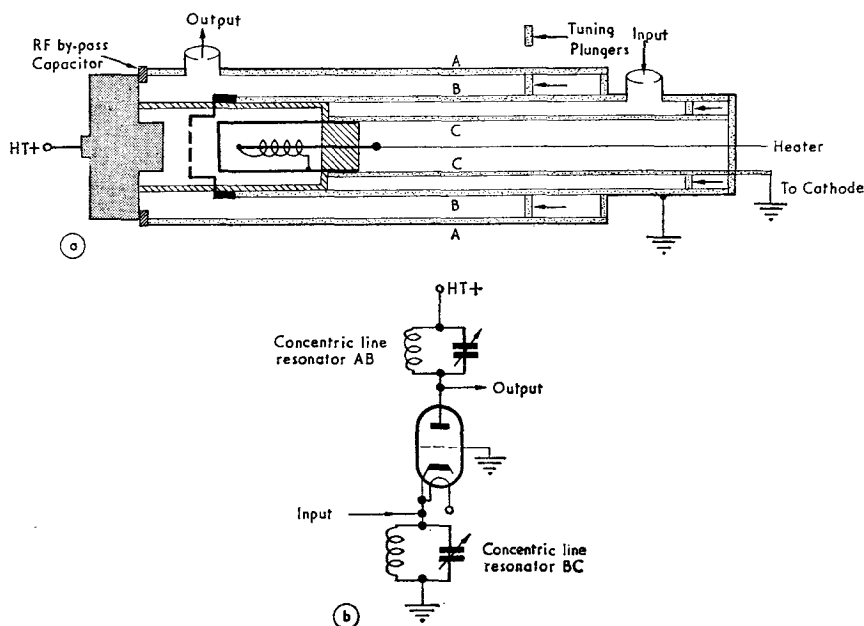


FIG 5. UHF CONCENTRIC LINE AMPLIFIER

a feedback probe between the two resonators. We then have positive feedback from output back to input. The *frequency* of operation is determined by the *length* of the coaxial resonators and this can be adjusted as required by varying the tuning plungers.

Planar triodes of the disc-seal type are manufactured over a wide range of power-handling capabilities. They may be used in the r.f. amplifier stage of a u.h.f. radar receiver where the signal power is of the order of microwatts. They are also designed for use in the final r.f. power amplifier stage of some u.h.f. radar transmitters where they must be capable of providing *average* power outputs of the order of 1 kW.

To obtain *large average* power outputs the valve must be capable of dissipating the heat developed at the electrodes. Very high power valves use water or air-blast cooling on grid and anode. The larger the valve, the more heat it can dissipate and the greater is the power output. However, because of transit time effects and other considerations, the *size* of the valve must be *proportional to wavelength*. Thus the *higher* the frequency, the *smaller* is the valve and the *smaller* is the power that the valve can handle. In addition, the *peak* power output of a valve in pulsed radar is limited by the electron emission available and by the breakdown voltage level between electrodes. Typical high-power planar triodes can be designed to produce an *average*

power output of about 1 kW at frequencies of the order of 1,000 Mc/s. *Peak* powers during pulses can be higher than 100 kW.

Beam power tetrodes have also been constructed as planar disc-seal valves, in some cases with the resonant cavities an integral part of the valve. The tetrode has a higher gain than the triode so that less driving power is required for a given output. In addition, the screen grid gives greater isolation between input and output and reduces internal feedback effects. On the other hand, the additional grid of the tetrode requires a more complex construction than the triode, and cooling becomes more of a problem.

One way of increasing the power output of grid-controlled valves at u.h.f. is to use a *number* of unit structures *in parallel* and to arrange them in a coaxial, cylindrical configuration, all within the same vacuum envelope (Fig 6). This form of multi-unit construction means that the valve can be connected directly to the resonant cavities. At the same time, with a number of units operating in parallel, a very high power output can be obtained since the heat to be dissipated is spread over a relatively large area.

To give some idea of the very high powers that such valves are capable of producing, the following figures are quoted. A multi-unit planar tetrode designed for u.h.f. radar produces an *average* power output of 10 kW and a *peak* pulse power output of 2 MW at a frequency of 400 Mc/s. It operates with a pulse duration of $10\mu\text{s}$ at a p.r.f. of 400 p.p.s. The valve measures only 8 in. diameter by 8 in. long and weighs 38lbs. It is water cooled. It has a gain of 20 db and operates with an efficiency of 55 per cent. The required anode voltage is 50 kV and the anode current during the pulse is 75 amperes.

Even *higher* values of average and peak power have been obtained. One multi-unit planar triode has been designed to produce 500 kW average power and 10 MW peak pulse power at a frequency of 500 Mc/s.

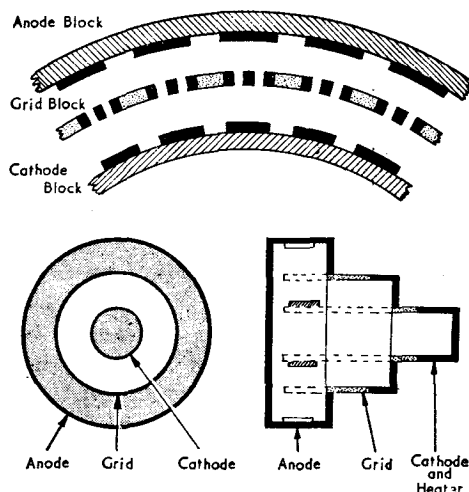


FIG 6. MULTI-UNIT PLANAR TRIODE

Microwave Valves Used at UHF

Almost all the valves which, up till now, have been considered as 'microwave' devices can be 'scaled down in frequency' to the 300-1,000 Mc/s range. Such valves include the magnetron oscillator, the multi-cavity klystron amplifier, the travelling-wave tube and the amplatron. The first three are mentioned briefly in Part 1B of these notes (p 491) and, together with the amplatron, they are considered in more detail in Section 4 of this book.

The *magnetron oscillator* provides a high *peak* power output and is used in radar transmitters which do not employ a r.f. power amplifier. It finds its main application in the microwave region and is not greatly used in the 300-1,000 Mc/s band. At u.h.f., the m.o.-p.a. type of transmitter is more usual because one of the reasons for using the u.h.f. band is to utilise transmitters capable of giving a high *average* r.f. power output. The magnetron cannot provide this. However, where economy and simplification in circuitry are required in u.h.f. transmitters, the magnetron may be used. A typical u.h.f. magnetron designed for use in a m.t.i. radar operates at a frequency of 430 Mc/s. For a $6\mu\text{s}$ pulse at a p.r.f. of 300 p.p.s. the peak power output is 2 MW, the average power being 3.6 kW.

With the requirement for radar transmitters of high average power, and the resulting movement towards the 300-1,000 Mc/s band to achieve this more easily, *multi-cavity klystrons* were developed for this frequency band. Most klystrons for the u.h.f. band have three or more resonant cavities. The advantages of having more than two cavities are an increase in gain, greater efficiency and wider bandwidths. One klystron designed for use in the u.h.f. band has a peak power output of 30 MW and an average power of 30 kW with a 10 μ s pulse. Even higher average powers of the order of 100 kW have been achieved. Note however that these very high power multi-cavity klystrons are also *very large*. One klystron rated at 75 kW average power is 10 ft. 6 in. high.

Travelling-wave tubes have also been designed as power amplifiers in the u.h.f. band. Their main advantage compared with the klystron is their larger bandwidth. Like the klystron, high-power travelling-wave tubes are very large devices. A travelling-wave tube designed to give a peak pulse power output of 100 MW at a frequency of 400 Mc/s is 18 ft long and weighs over 4 tons. They are not normally as large as this however. Peak power outputs of the order of a few MW are more usual.

The *amplitron* is an amplifier, similar in many ways to the magnetron oscillator. It is considered in more detail in Section 4. Amplitrons provide high peak and high average values of power output, they have a wide bandwidth, and they operate at a high efficiency. But, compared with the klystron and travelling-wave tube, they have a *low gain*. UHF amplitrons are capable of 5 to 10 MW of peak power output over a 10 per cent bandwidth, at a conversion efficiency of about 90 per cent. The very high efficiency means that amplitrons are able to handle large powers with structures of reasonable size.

Comparison of UHF High-power Valves

The klystron, travelling-wave tube and amplitron amplifiers can be designed to operate anywhere within the normal radar frequency range from 300 Mc/s upwards. The grid-controlled valve is usually confined to operation at u.h.f. or lower. Below 1,000 Mc/s the grid-controlled valve is capable of delivering more average power than any of the others mentioned. Its physical size is usually also much less. However, the power output of any valve falls off with increase in frequency, and at frequencies above about 1,000 Mc/s the grid-controlled valve is seldom used. The 'microwave' devices are better at frequencies above u.h.f. All the amplifier valves considered are capable of generating high average powers of the order of tens of kilowatts at u.h.f. For comparison, a good average power for a magnetron oscillator is a few kilowatts.

UHF Radar Receivers

We noted in Section 1 (p 31) that a radar receiver must have *high sensitivity* and *high gain* in order to receive very weak echo signals from long range; it must have a *wide bandwidth* to preserve the shape of the received pulse; and it must have a *low noise factor* so that the signal-to-noise ratio is kept at a high value to prevent the signal being masked by noise generated within the receiver itself. We shall deal with all these points in much more detail in Section 6 when we consider radar receivers.

It is difficult to obtain a low noise factor and, as we noted in Section 1, it is this which, at the moment, is the main limitation to increasing radar range. In the past, most radar receivers have dispensed with a r.f. amplifier stage because conventional r.f. amplifiers are too noisy and tend to decrease the signal-to-noise ratio. In such receivers the crystal mixer is the first stage. However, progress in the development of low-noise amplifiers is such that many radar receivers now use a r.f. amplifier stage to improve the signal-to-noise ratio. The low-noise amplifiers which have been used include the low-noise planar triode, the travelling-wave tube, the parametric

amplifier, the maser and the tunnel diode amplifier. With the exception of the low-noise triode, the amplifiers mentioned are usually considered as 'microwave' devices and they are dealt with as such in more detail in Section 4. However, as we have seen, many microwave devices may be scaled down in frequency for use at u.h.f. and in the following paragraphs we compare those devices which have been used as r.f. amplifiers in u.h.f. radar receivers.

A graph showing how the noise generated by an amplifier varies with frequency for various types of amplifier is shown in Fig 7. The noise level of a crystal-mixer (i.e. receiver with no r.f. amplifier) may be taken as the reference.

From this graph it is seen that a low-noise *disc-seal triode* gives an improvement over a crystal mixer at frequencies below about 1,000 Mc/s. The triode will normally be operated in a grounded-grid circuit. An additional advantage of using a triode is that it is less liable to burn-out than a crystal mixer.

The *parametric amplifier* introduces even less noise in the u.h.f. band and has been used in 50 cm u.h.f. radar receivers. It is claimed that by using a parametric amplifier a 100 per cent increase in range is obtained relative to the same system using a crystal-mixer receiver and a 40 per cent increase relative to the same system using a low-noise disc-seal triode. The parametric amplifier is also used in the microwave region although, like most devices, the noise which it introduces increases with frequency.

UHF masers have also been developed. At present they have the lowest noise figure of all amplifiers. However, at frequencies below 1,000 Mc/s there is little to choose between the noise level of parametric amplifiers and that of masers; because of the complexity of the maser, parametric amplifiers are generally preferred at the lower radar frequencies. As we shall see when we discuss these devices in Section 4, the maser comes into its own in the microwave region.

Travelling-wave tube amplifiers introduce less noise than a crystal mixer. They are not as good as the parametric amplifier at the lower radar frequencies and so are seldom used at u.h.f. However towards the X band (10,000 Mc/s) there is little to choose between them.

The *tunnel diode* is better than a crystal mixer but not as good as the other low-noise devices. It is not used at u.h.f. but it may find application in microwave systems where its simplicity and lower cost are bigger considerations than its relatively high noise level.

Summary

In this chapter we have seen that u.h.f. radar, despite its limitations, has many useful fields of operation, especially in long-range and c.w. radar systems. We have touched very briefly upon many devices which are new to us. These are considered more fully in Section 4, and their effectiveness will become more apparent then. Among the problems still waiting to be examined are the aerial systems and scanning methods used in u.h.f. radar. These are discussed in the next chapter.

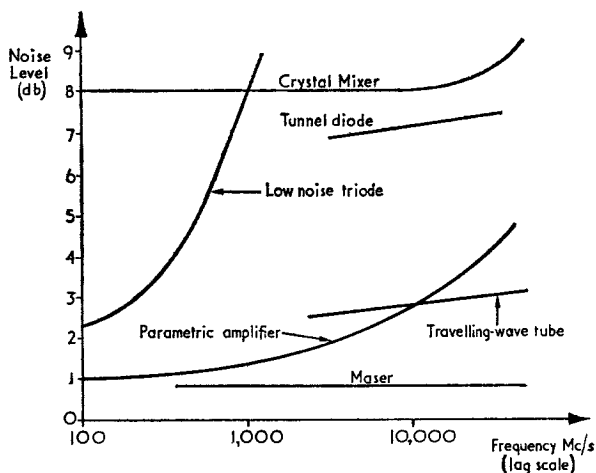


FIG 7. NOISE LEVELS OF UHF AMPLIFIERS

CHAPTER 2

UHF RADAR AERIAL SYSTEMS

Introduction

We have noted in earlier chapters of this book that a radar aerial system is normally made large in terms of the wavelength being radiated. There are two main reasons for this:—

a. The greater the dimensions (or *aperture*) of an aerial system in terms of wavelength the narrower is the beamwidth of the radiated energy. For accurate angular measurement and discrimination a narrow beamwidth is essential.

b. The greater the aerial aperture the greater is the amount of echo energy collected by the aerial, and hence the greater is the input signal to the receiver.

For S band and X band radars, and for those working at even higher radar frequencies, the very short wavelengths mean that relatively small aerial systems (though large in terms of wavelength) can be used. Parabolic reflector types of aerial, and similar systems, are convenient at these high radar frequencies. They can be made to produce the required narrow beamwidths fairly easily and they are also easy to move mechanically for scanning. Aerials for these high radar frequencies are considered in Section 4.

As we have seen, however, aerial systems for the lower radar frequencies (below 1,000 Mc/s) tend to be physically huge structures, and reflector type aerials then become less practicable for reasons we shall see shortly. The type of aerial system best suited to the lower radar frequencies is the *aerial array*. An aerial array consists of a number of individual aerial elements (e.g. half-wave dipoles) so spaced and excited as to produce a narrow beam in a required direction. The *broadside array* considered in Part 1B (p 465) is an example of an aerial array.

Advantages of Aerial Array at UHF

We have seen that large aerial apertures are necessary for high-performance, long-range radars. Reflector-type aerials for use at u.h.f. would be difficult to construct and would be costly. The aerial array is relatively easy to construct and, because of the longer wavelength at u.h.f., each element in the array would be large compared with an element designed for centimetric radars.

A half-wave dipole at 600 Mc/s is about 1 ft long; at 3,000 Mc/s it is only about 2 in. Thus the *number* of elements needed to fill the aerial aperture at u.h.f. is fairly small.

The aerial array may be used as a fixed-beam aerial, the beam being scanned in azimuth by mechanical rotation of the entire structure. This is the type of aerial system used in the metric radar type 7 discussed in Section 1 (p 44). An illustration of part of the aerial array used is shown in Fig 1 overleaf. It is a stacked broadside array, containing eight stacks of full-wave dipoles with four dipoles in each stack. It produces a beamwidth of about 5° in azimuth and 10° in elevation. The reflector screen, spaced a few inches behind the dipoles to ensure radiation in a forward direction only, measures 64 ft by 10 ft. A parabolic reflector type of aerial system for the same frequency (220 Mc/s) would be a less compact structure. Fig 1 shows that the aerial array is relatively 'flat', whereas a parabolic reflector would need considerable 'depth', as well as large cross-sectional area, to support a feed at the focal point of the reflector.

The aerial array also has the ability to 'steer' a beam 'electronically' without having to move large mechanical structures. This is an important advantage if the aerial array is large, as it will be at u.h.f. The aerial array has this inherent ability because the individual aerial elements can each be controlled independently.

Other advantages of the aerial array are that *more than one* beam can be generated within the same aerial aperture, and large peak powers can be radiated.

In this chapter we shall take a closer look at the aerial array, with particular reference to how the beam is electronically scanned and how stacked beams are produced.

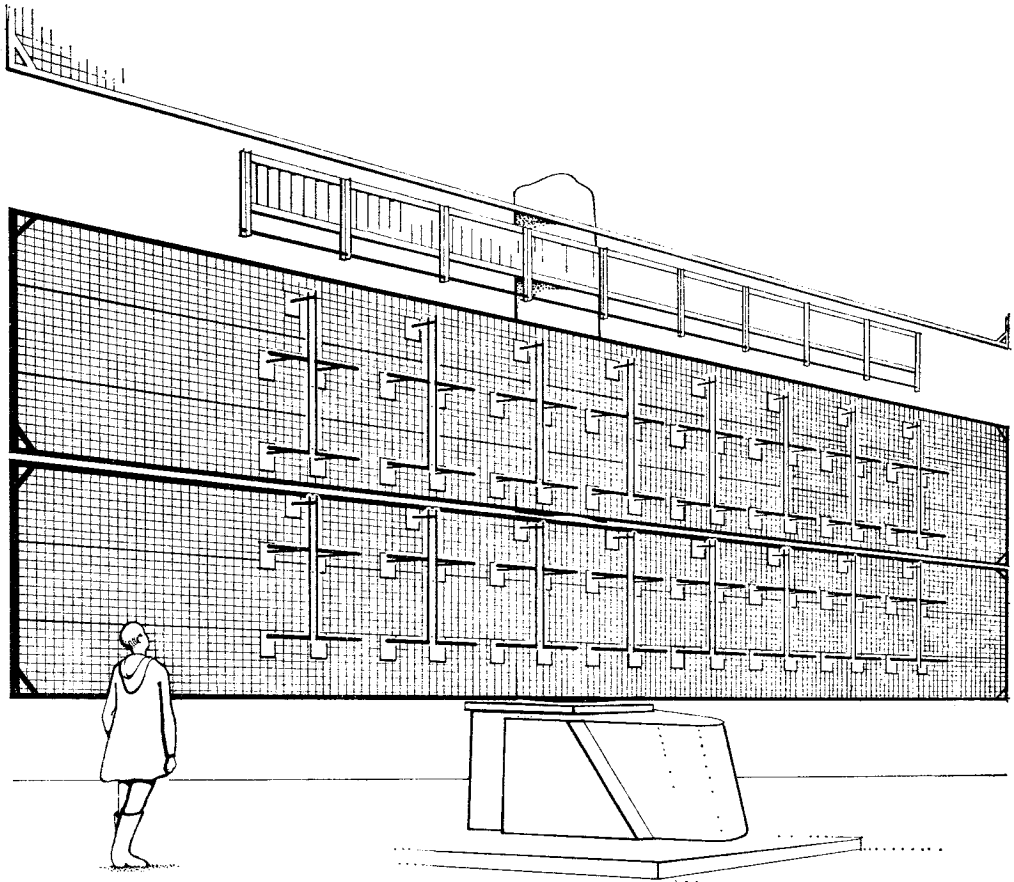


FIG 1. MECHANICALLY-SCANNED AERIAL ARRAY

Electronic Scanning

The beam of an aerial array may be steered in space by electronic means by varying the *phase* of the input to each element in the array. In the broadside array discussed in Part 1B, each element is fed *in phase* with equal amplitude signals. The power radiated by each element then combines in such a way that the beam produced by the array is *broadside* to the line of the array (Fig 2a).

By introducing a *phase difference* between the input to each element the power due to each element combines in such a way that the main beam now points in a direction *other than broadside* (Fig 2b). This principle is demonstrated in the *end-fire* array, considered in Part 1B (p 468), where the phase difference between each element is such that the beam is produced *in line* with the line of the array.

In addition, it may be seen that if the phase difference between adjacent elements is made

variable, the beam may be *steered* as the phase is changed. Practical values of the angle θ in Fig 2b for electronic scanning are $\pm 30^\circ$.

Fig 2 could be taken to represent the horizontal radiation pattern of a row of vertical dipoles. Fig 2b then illustrates scanning *in azimuth*. Alternatively, Fig 2 could be taken to represent the vertical radiation pattern of a tier, or stack, of horizontal dipoles, one above the other. In this case Fig 2b represents scanning *in elevation*. Thus, electronic scanning is possible both in azimuth and in elevation. In practice, to produce a narrow beam in *both* dimensions, a stacked broadside array may be used, as in Fig 1. With this type of array, we shall see that it is possible to have simultaneous scanning in azimuth and in elevation.

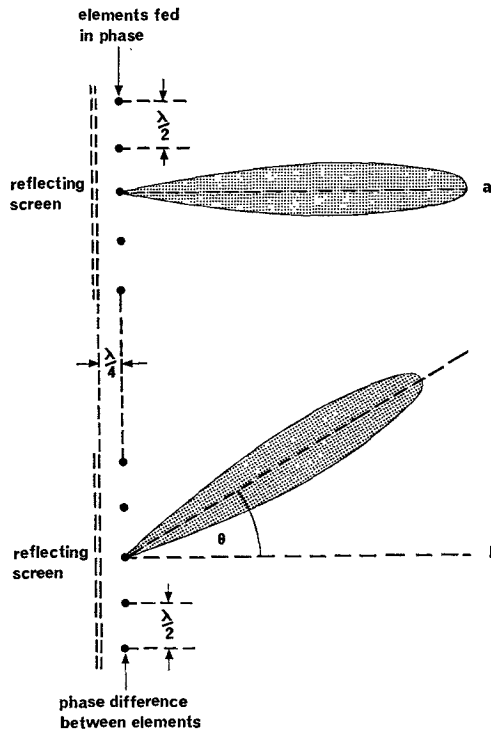


FIG. 2. PRINCIPLE OF ELECTRONIC SCANNING BY VARYING THE PHASE TO THE ELEMENTS

Methods of Feeding Array

The variable phase shift necessary to produce electronic scanning may be obtained with either a *series-fed* or a *parallel-fed* arrangement. Fig 3a shows a series-fed arrangement in which the energy is transmitted from one end of the line. Between each element we have identical phase-shifters, each of which introduce the same phase shift ϕ between adjacent elements. Thus, taking element 1 as reference (zero phase), element 2 has a phase of ϕ degrees, element 3 has a phase of 2ϕ degrees, and so on. The direction in which the resulting beam points depends upon the value of ϕ . We have seen that when ϕ is zero, the beam is broadside to the array. For positive values of ϕ the beam swings more and more to one side as ϕ is increased. For negative values of ϕ the beam swings to the other side. Thus by continuously varying ϕ backwards and forwards between given positive and negative values we cause the beam to scan.

Fig 3b shows a parallel-fed arrangement. The energy to be radiated is divided between the elements. Equal lengths of line transmit the energy to each element so that no uncontrolled phase shift is introduced by the lines. Again there is a phase shift of ϕ between adjacent elements and relative to element 1, of ϕ , 2ϕ , 3ϕ , etc. for elements 2, 3, 4 etc.

The series-fed arrangement introduces more loss than the parallel-fed system because the phase-shifters are all connected in series. However, the series-fed arrangement has the advantage that, since the phase shift introduced by the phase-shifters must be the *same* at any given instant, a *single* control signal varying *all* the phase-shifters by the same amount is sufficient to steer the beam. In the parallel-fed arrangement, *each* element (apart from element 1) has to be controlled separately.

Resonant and Non-resonant Arrays

A *resonant* broadside array is one in which the elements are spaced exactly a half wavelength apart. When each element is fed in phase, the array then radiates a beam broadside to the

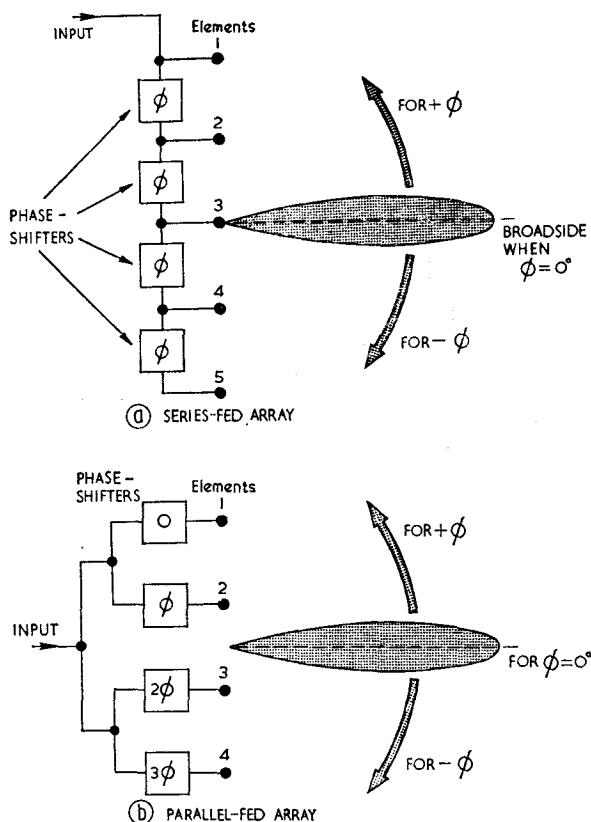


FIG 3. SERIES AND PARALLEL METHODS OF BEAM-STEERING

line of the array. However, like all frequency-sensitive devices, it has a *narrow bandwidth*. If the input frequency changes, the elements are no longer a half wavelength apart. Matching between the array and the line is then no longer correct; standing waves are set up in the lines and the power radiated falls off. In addition, the beam is no longer radiated broadside to the

line of the array. A *squint* is introduced (see Section 4 on slotted waveguide array). The squint results from the fact that the change in frequency has caused the lines joining adjacent elements in the array to be longer or shorter than a half wavelength (depending upon whether the frequency has decreased or increased). This introduces a phase shift between elements and the beam is steered slightly to one side or the other.

For some applications it is better to make the array *non-resonant* by spacing the elements greater or less than a half wavelength apart at the centre operating frequency. Although the beam now has a squint, the non-resonant array has a *wide bandwidth* and this may be a requirement.

Phase-shifting Devices

Variable phase-shifters are usually based on changing the *electrical length* of a transmission line (see Part 1B p 447). This may be done mechanically by physically shortening or lengthening the line, or it may be done electronically by changing the electrical length of the line directly.

Various forms of *mechanically-controlled* phase-shifters have been used in radar equipments. Some of these are more applicable to microwave systems; others can operate down to u.h.f. One of the simplest, suitable for use at u.h.f., is illustrated in Fig 4. It consists of a length of open-wire transmission line with a telescopic section whose length can be varied. This type of phase-shifter is referred to as a *line stretcher*. As the U-shaped telescopic section is extended (rather in the manner of a trombone slide) the total length of the line from input to output terminals is varied. This causes a variation in the phase of the output signal. With suitably controlled phase-shifters of this kind between each element of the array we can produce electronic scanning. Other types of phase-shifter have been produced which also change the physical lengths of a transmission line connected between the input and the aerial element. The principle is the same as that described above for the line stretcher and will not be considered further here.

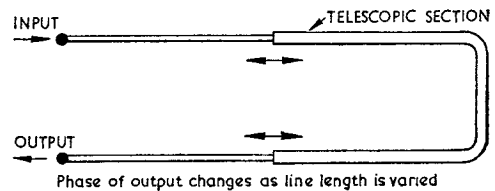


Fig 4. LINE-STRETCHER AS A PHASE-SHIFTER

If the aerial array is fed from a waveguide, a change in the *dimension* of the waveguide varies the phase of the signal passing through it. Thus by mechanically varying the waveguide dimensions we can produce electronic scanning of a beam. This is considered in more detail in Section 4.

The time required to vary the phase of a phase-shifter through 360° depends upon the type of phase-shifter. For mechanically-controlled phase-shifters, switching times of 0.1 second are possible. Such speeds allow the aerial beam to be scanned much faster than is possible with an aerial array which is moved mechanically.

Even faster switching times, of the order of microseconds, may be obtained by using *electronically-controlled* phase-shifters. However, very fast scanning times may not be a requirement. It was pointed out in Section 1 (p 36) that the scanning speed must be related to the p.r.f. of a pulsed radar in such a way that at least one pulse per scan illuminates the target. For c.w. radar, of course, there is no such limitation in scanning speed.

Among the electronically-controlled phase-shifters which have been used in practice are *ferrite phase-shifters* and *travelling-wave tubes*.

A ferrite phase-shifter is a length of transmission line, or waveguide, in which a strip of ferrite material has been placed. The phase of the output signal may be varied by changing the *magnitude* of the d.c. magnetic field applied to the ferrite. This principle is dealt with in Section 4, Chapter 6. Phase shifts of 360° can be obtained with relatively small d.c. magnetic field variations.

The travelling-wave tube can be made to provide a fast, electronically-controlled phase shift by varying the voltage applied to the helix. This varies the velocity with which the energy is propagated and so varies the phase. Only a small voltage variation is needed to obtain the necessary phase-shift. An advantage of the travelling-wave tube as a phase-shifter is that it can also provide amplification of the signal over a wide bandwidth. It cannot, however, be used in both transmitting and receiving roles. Its main use is in aerial arrays designed for reception (see p 246).

Other devices, including special circuit arrangements, may be used to provide an electronically-controlled phase shift. These will not be considered here. They are more appropriate to Section 4, where they are dealt with after the theory of waveguides, hybrid junctions, etc. has been covered. Enough has been said to show the various possibilities.

Frequency Scanning

We saw on p 447 of Part 1B that the *electrical* length of a transmission line depends upon the *frequency* of the signal applied to it. Suppose a given length of transmission line is exactly one wavelength long at a frequency f_0 . Then at the frequency f_0 , the output is *in phase* with the input (Fig 5). If the frequency is *increased*, the electrical length of the fixed line becomes *greater* than a wavelength. The line now behaves as a reactance and the output is *out of phase* with and *leads* the input by an amount that depends upon the frequency deviation above f_0 . Conversely, if the frequency is *decreased*, the electrical length of the fixed line behaves as a reactance of *opposite* sign compared with that due to an increase in frequency above f_0 . The output is again *out of phase* with the input, and *lags* it by an amount that depends upon the frequency deviation below f_0 . Thus, by using *given lengths* of transmission line and varying the frequency of the input above and below f_0 we can cause the *phase* of the signal applied to the aerial elements to vary.

A basic arrangement is illustrated in Fig 6. The lines connecting adjacent elements of the array are of equal length and so chosen that the phase at each element is the *same* when the input is at the frequency f_0 . The resultant beam is radiated *broadside* to the array. As the input frequency is increased above f_0 , the phase shift through each section of transmission line *increases by the same amount*. If the phase shift increases from 0° to $+\theta^\circ$ we then have a phase shift of θ° between adjacent elements and of $\theta, 2\theta, 3\theta$, etc. relative to element 1 for elements 2, 3, 4, etc. This causes the beam to be steered to one side. For frequencies below f_0 the phase shift is of *opposite* sign and the beam is steered to the other side. Thus, as the input frequency is swept either side of f_0 , the beam is scanned over a given sector.

The main disadvantage of frequency scanning is that a relatively large bandwidth is required.

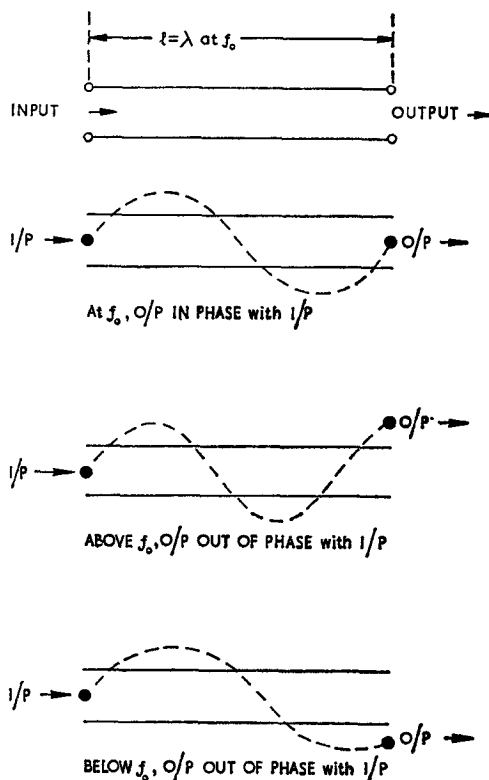


FIG 5. EFFECT OF FREQUENCY ON ELECTRICAL LENGTH OF LINE

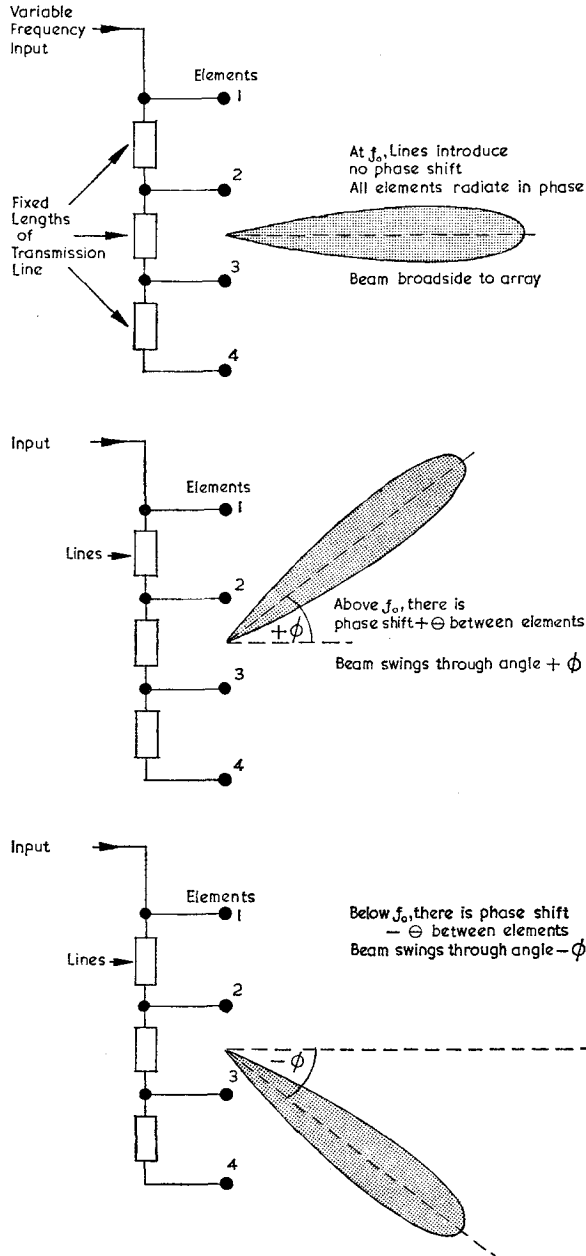


FIG 6. FREQUENCY SCANNING

Scanning in Azimuth and in Elevation

We have seen that it is possible to steer a beam *in azimuth* by varying the phase shift between adjacent elements in a linear broadside array. Alternatively, if the elements are arrayed in a vertical tier, or stack, then varying the phase shift between adjacent elements will steer the beam

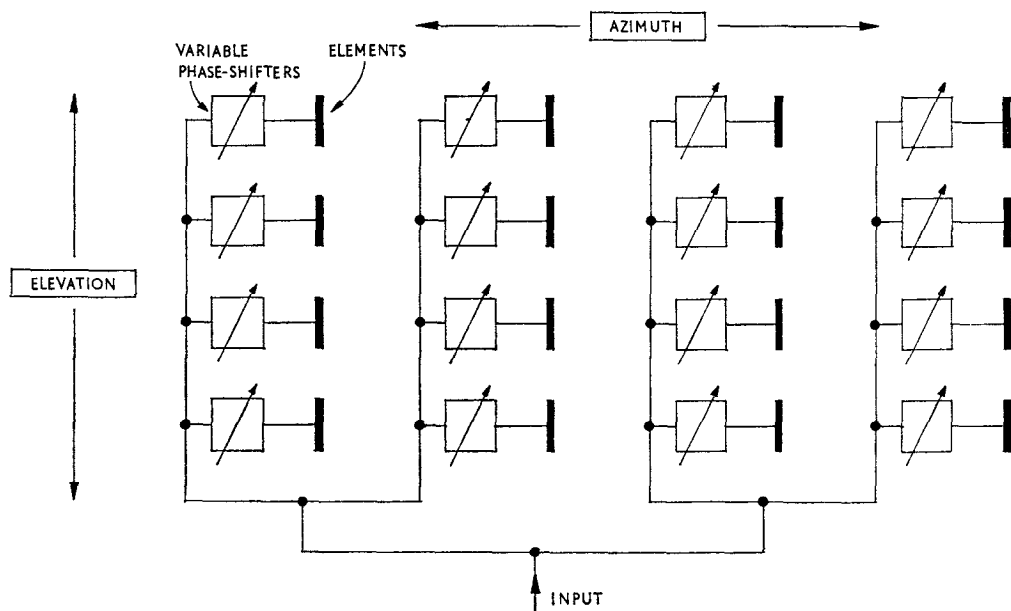


FIG 7. ELECTRONIC SCANNING IN BOTH AZIMUTH AND ELEVATION

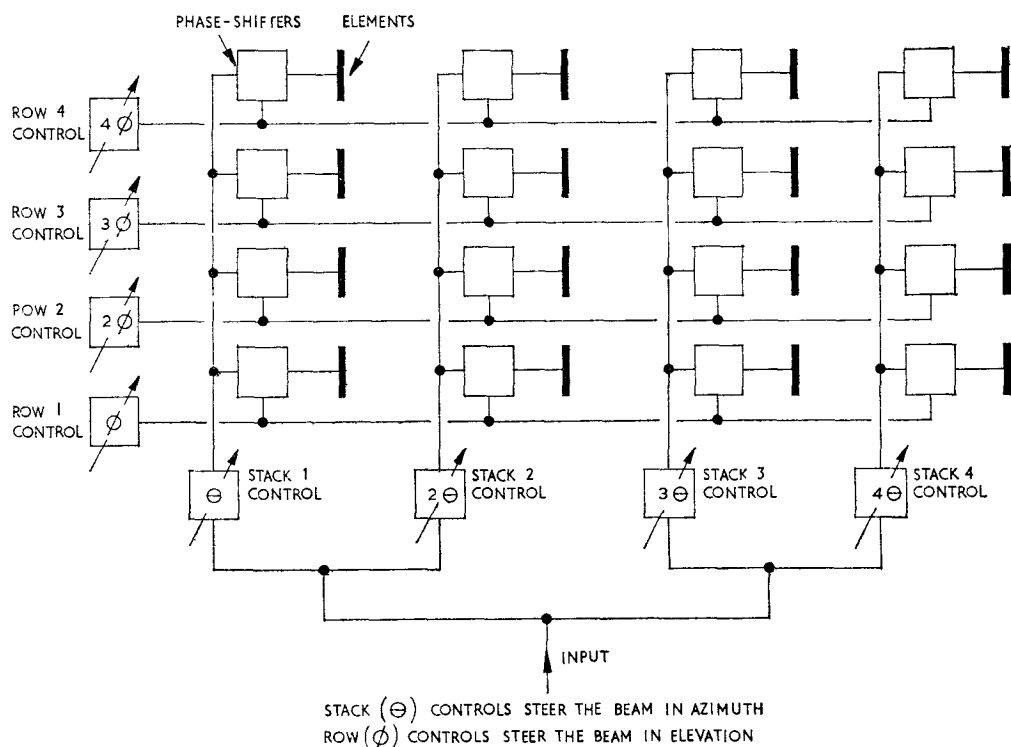


FIG 8. REDUCING THE CONTROLS FOR SCANNING IN TWO DIMENSIONS

in elevation. It is, of course, possible to combine these two systems by having a *stacked broadside array* arranged to give scanning both in azimuth and in elevation. Fig 7 illustrates such an array. It consists of four stacks of elements arranged in a linear broadside array with four elements in each stack. In practice, many more elements would be used. The beam generated by this two-dimensional array is scanned by applying to each element the phase shift necessary to position the beam in the required direction. The correct phase for each element is found by *superimposing* the phase shift needed to position the beam at the required angle in azimuth with that required to position the beam in elevation. Varying this *composite* phase shift in the correct manner will cause the beam to scan in both dimensions. With this system each element has to be controlled *independently* so that there are as many controls as there are elements.

We can reduce the number of controls needed by arranging to steer the beam *independently* in azimuth and in elevation (Fig 8). In this array all the elements which lie in the first vertical stack receive the same phase shift θ ; the second vertical stack has phase shift 2θ ; and so on for the third and fourth stacks. Thus, between each stack we have phase shift θ , and as θ is varied so the beam scans *in azimuth*.

Similarly, all the elements which lie in the first horizontal row receive the same phase shift φ ; the second horizontal row has phase shift 2φ ; and so on for the third and fourth rows. Thus, between each row we have phase shift φ , and as φ is varied so the beam scans *in elevation*.

Compared with the first system illustrated in Fig 7 when 16 control signals were needed, the system of Fig 8 requires only 8 control signals. In a practical array containing many more elements the advantages of the second system are even more obvious.

A combination of *frequency scanning* in one plane and phase-shifters for scanning in the other plane is also possible. Another possibility is electronic scanning in elevation combined with mechanical rotation of the array for scanning in azimuth. All the systems mentioned have been used in practice.

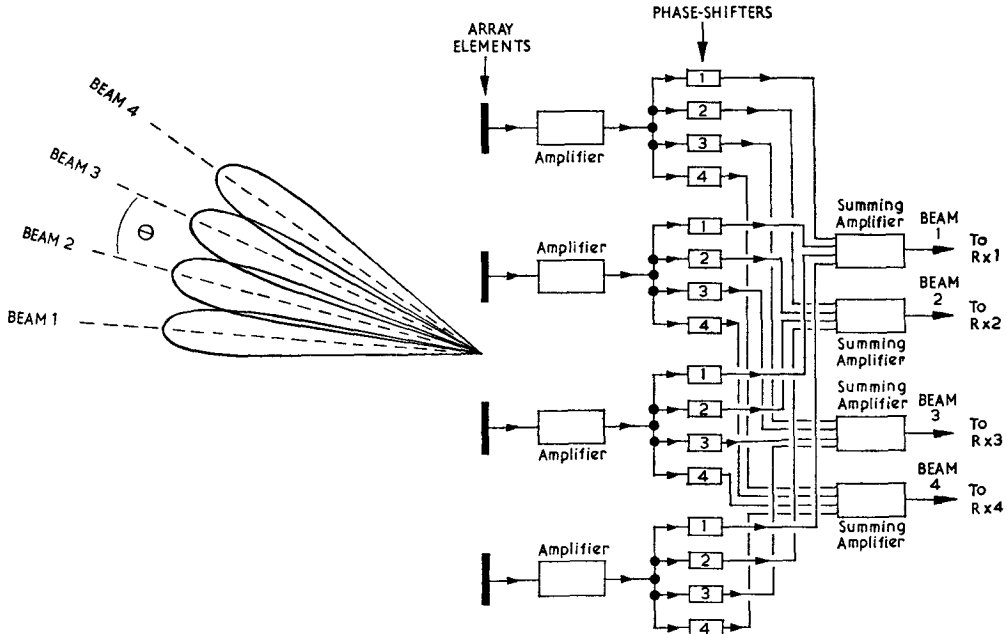


FIG 9. BASIC IDEA OF MULTIPLE BEAM PRODUCTION

Multiple Beam Arrays

It was mentioned earlier in this chapter that an array is capable of generating a *number of beams* simultaneously from the same aperture. Because of the lower power levels, this is usually easier to achieve in an array designed for reception than in one designed for transmission, and it is normally as a receiving aerial that the multiple beam array is used.

The simple stacked receiving array which produces a single beam in elevation can be converted to a multiple beam aerial by inserting additional phase-shifters at the output of each element. Each beam to be formed requires one phase-shifter to each element. Thus to produce four separate beams in elevation we need four phase-shifters for each element as shown in Fig 9.

The output from each element is amplified and then sub-divided for application to the phase-shifting (beam-forming) networks. Corresponding phase-shifters in each element introduce the same phase shift. The outputs from phase-shifter 1 in each element are combined in summing amplifier 1 to provide beam 1. Beams 2, 3 and 4 are formed in a similar way. The signal due to each separate beam is then applied to separate receivers for processing. Thus each receiver is concerned only with those signals which originate from targets within a particular sector in

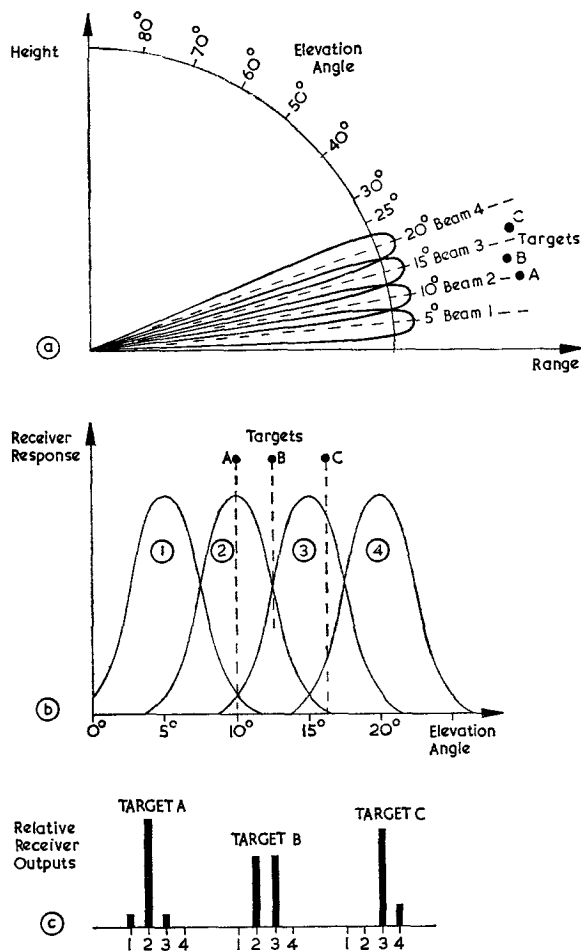


FIG 10. USE OF MULTIPLE BEAMS IN HEIGHT-FINDING

elevation. A beamwidth of 1° in elevation and 5° in azimuth is common for each beam. The angle θ between adjacent beams depends upon the phase shift introduced by the phase-shifters and this is normally fixed to give a small overlap of adjacent beams.

This type of array finds application in *height-finding* radars. The target is normally 'illuminated' by a conventional search radar which provides information on range and bearing of the target. A normal fan-shaped beam will be used for this. To find the *height* of the target the multiple beam receiving array is rotated until it is aligned on the target heading, all the beams being *fixed in elevation*.

In Fig 10a target A is seen to be directly in line with beam 2, so receiver 2 gives maximum output, the much smaller outputs from receivers 1 and 3 being equal. This indicates that target A has the same elevation angle as beam 2, namely 10° . Target B lies between beams 2 and 3, its position being such that the outputs from receivers 2 and 3 are equal; receivers 1 and 4 give no output. Thus target B is equidistant from the centre lines of beams 2 and 3 and has an elevation angle of 12.5° . The elevation angle of target C can be calculated by interpolating the separate receiver outputs. In practice, accuracies of the order of 0.1° can be obtained. This is equivalent to a height resolution of less than 1,000 ft at a range of 50 miles. The more conventional radars using a beaver-tail beam for height-finding (see p 34) cannot give a resolution better than several thousand feet at this range.

A practical equipment which uses this system of height-finding generates over 100 separate beams in elevation. The lowest beam is at an angle of 0.5° above the horizontal and the highest beam is at an elevation angle of 40° . The beamwidth is 1° in elevation and 5° in azimuth. The accuracy is 0.1° .

In this chapter we have considered electronic scanning and the production of stacked beams in relation to u.h.f. aerial arrays. These systems are, however, also possible in centimetric radar where reflector type aerial systems are more usual. Electronic scanning, as applied to microwave aerial systems, is considered in Section 4.

SECTION 4

CENTIMETRIC RADAR

Chapter	Page
1 Resonant Cavities 	249
2 Klystrons 	255
3 Travelling-wave Tubes 	263
4 Magnetrons 	269
5 Waveguides 	279
6 Waveguide Components 	289
7 TR Switching and Frequency Changing 	305
8 Centimetric Aerials 	317
9 Microwave Semiconductor Devices 	341
10 Masers	355

CHAPTER 1

RESONANT CAVITIES

Introduction

Radars which operate on metric wavelengths have certain limitations. A fairly large aerial system is necessary to obtain a reasonably narrow beam; the high peak powers which are required for long ranges are difficult to obtain; and a narrow pulse duration with its advantages of short minimum range and good target definition cannot easily be produced at metric wavelengths.

These limitations led to the development of components which operate at *centimetric* wavelengths. These devices employ different techniques and are of different construction from those used at longer wavelengths; they are known as *microwave* components and are designed to operate at frequencies above 1 Gc/s (1,000 Mc/s). Most microwave radars work at either 10 cm (3 Gc/s) or 3 cm (10 Gc/s) and are referred to as centimetric radars. Some meteorological radars operate at shorter wavelengths (higher frequencies) in the millimetric region.

The advantages obtained by using centimetric radars may be summarised as follows:

- a. Small, light, compact aerials with narrow beam widths can be designed to give more accurate bearings, stronger echoes and less chance of interference with other transmissions.
- b. Because of the higher frequency, more r.f. cycles can be included for a given pulse duration and steep-sided short-duration pulses can be produced. These give more accurate range measurements and better target definition. Short pulses mean that a higher peak power for a given mean power can be handled by the components.
- c. The bandwidth of the radar receiver can be increased, giving better pulse reproduction.

The design of centimetric transmitters and receivers involved many problems which were solved by new techniques. Some of the problems and the devices invented to overcome them are mentioned below. Details of the construction and principles on which these devices work form the subject of this and succeeding sections.

At centimetric wavelengths open wire and coaxial lines are inefficient because of radiation and resistance losses. Short lengths of coaxial cable may be used but the main lengths of transmission line between transmitter and aerial and receiver and aerial are *waveguides* which at centimetric wavelengths are of conveniently small cross-section. Even more compact lines using microwave printed circuit techniques are used in some radars.

To produce well-shaped pulses as short as $0.1\mu\text{s}$ new techniques in *pulse modulation* were devised. These use pulse-forming networks and soft valve switches to provide the high modulating powers required.

Conventional valves or transistors cannot be used as amplifiers or oscillators at centimetric wavelengths because of the effects of transit time. Some microwave amplifiers and oscillators employ techniques which make use of electron transit time. Examples of these devices are the *klystron*, the *magnetron* and the *travelling-wave tube* (t.w.t.). The low-noise parametric amplifier and the maser are other devices recently developed for amplification of microwaves. The tuned circuits used with these devices are built into the device, making the construction entirely different from the conventional amplifier or oscillator circuit.

Attenuation of an e.m. wave increases as its frequency is increased. In centimetric radar this is counterbalanced by using narrow beams which concentrate high energies on to the target.

Conventional tuned circuits using coils and capacitors cannot be used at centimetric wavelengths since their physical size would be impractical and losses due to radiation and resistance would be excessive. Hollow metal cylinders and cubes called *cavity resonators* or *resonant cavities* are used instead.

Resonant Cavities

At u.h.f. a short-circuited section of transmission line called a *lecher bar* may be used as the tuned circuit of an oscillator. Fig. 1a shows the instantaneous electric and magnetic fields and the current in an oscillating lecher bar. The direction of the current and of the fields changes each half cycle. If several lecher bars are connected in parallel (Fig. 1b) the inductance is reduced by the same amount as the capacitance is increased and therefore the resonant frequency ($f_o \propto \sqrt{\frac{1}{LC}}$) remains the same as for one lecher bar. However, the total resistance has decreased.

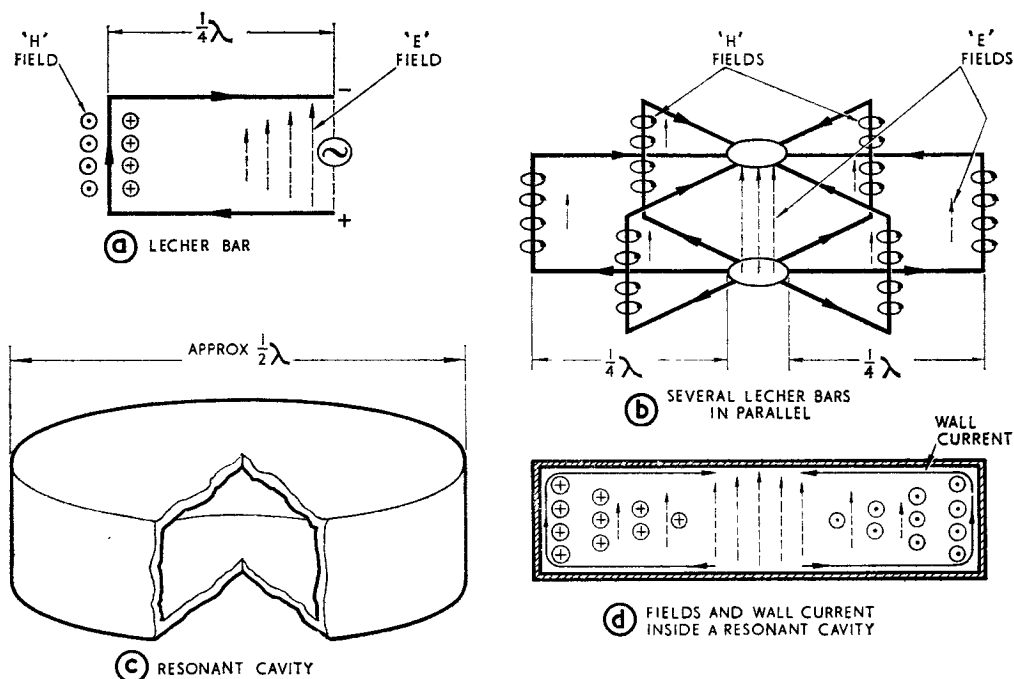


FIG. 1 DEVELOPMENT OF A RESONANT CAVITY

If an infinite number of lecher bars are placed in parallel a hollow cylinder or *cavity* is formed (Fig. 1c). Little radiation can occur and as the skin resistance is small total losses are small resulting in a very high magnification factor (Q).

The internal diameter of the cavity, which should be approximately half a wavelength, determines its resonant frequency. At centimetric wavelengths the cavity dimensions are small enough for it to form an integral part of the amplifier or oscillator valve.

The instantaneous electric (E) and magnetic (H) lines of force present when a lecher bar oscillates are shown in Fig. 1; they are alternating at a high frequency. The E lines may be associated with the circuit *capacitance* and the H lines with the circuit *inductance*. In a resonant cavity formed from an infinite number of lecher bars the individual H lines round each lecher bar combine as in a solenoid and form closed loops, strongest around the circumference of the cavity and weakening to zero at the cavity centre. This H field always lies parallel to the cavity walls. The E lines combine to form a strong E field at the centre of the cavity, decreasing to zero at the cavity walls. During each half cycle the E field builds up to a maximum at the instant that the H field falls to zero. During the next half cycle the opposite happens, i.e. the E and H fields are 90° out of phase as are the energies associated with the capacitance and inductance of a conventional tuned circuit.

The fields shown in the cavity cross-section of Fig. 1d are the fields *at one instant of time*. The e.m. energy in the cavity is oscillating at the cavity resonant frequency, the E and H fields changing direction every half cycle. The changing H field induces currents in the cavity walls. These *wall currents* are of the same frequency as the H field and always flow at right angles to the H field. Due to skin effect the currents flow only on the inner surfaces of the cavity. The outer surfaces can therefore be earthed without affecting the cavity operation.

The resonant cavity need not be cylindrical. Depending on the function of the cavity it may be rectangular, spherical or a modified cylindrical shape. In these cases the field pattern or *mode* which is formed inside the cavity when it is excited into oscillation will differ from that shown in Fig. 1. Nevertheless, whatever mode is formed, the E field will always be zero at the cavity walls and the E and H fields will always be at right angles to each other and 90° out of phase in time.

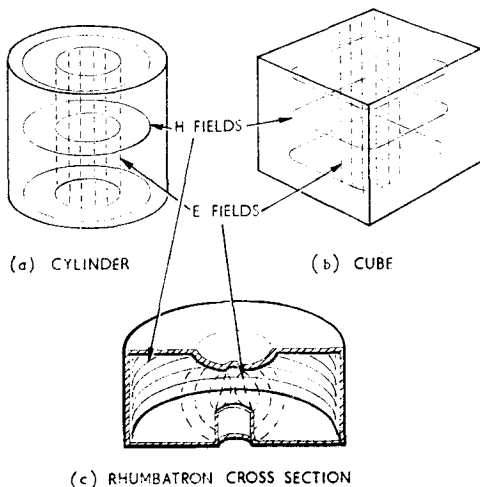


FIG. 2 COMMON CAVITY MODES

Some common cavity modes are shown in Fig. 2. The rhombatron cavity (Fig. 2c) forms the resonant element in a klystron valve. Electrons are passed through holes in the roof and floor of a cylindrical cavity and in order that the transit time of the electrons passing through the E field shall be small the shape of the cavity is modified as shown.

Cavity Coupling

In order to start oscillations in a cavity some method of coupling energy into the cavity must be provided. Also, once the cavity is oscillating some means of extracting r.f. energy is required. Perhaps the most commonly used method of coupling energy into or out of a resonant cavity is by means of a *loop coupler*. As shown in Fig. 3 the centre conductor of a short length of coaxial cable is extended beyond the outer conductor and formed into a loop. The loop is introduced into the cavity

at a position of *maximum H field* and for maximum coupling the loop should be at right angles to the H lines. The oscillatory H field then threads the loop inducing currents into it and energy may be extracted via the coaxial cable. Conversely, if the cavity is not oscillating it can be made to do so by feeding oscillatory currents at the cavity resonant frequency to the loop via the coaxial cable.

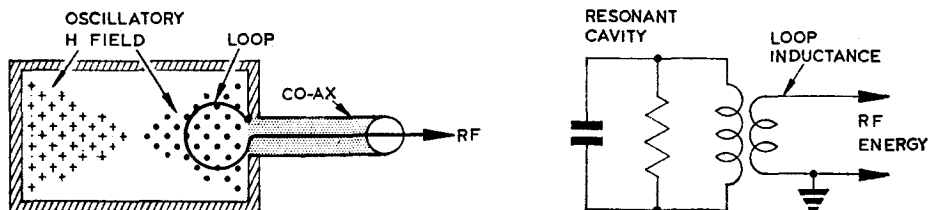


FIG. 3 LOOP COUPLING AND EQUIVALENT CIRCUIT

Another method of cavity coupling is shown in Fig. 4. The probe is an extension of the centre conductor of a coaxial cable and is inserted into the cavity at a position of *maximum E field* and parallel to the E lines. The oscillatory E field will induce voltages into the probe and energy may be extracted. Energy may also be fed into the cavity, the probe being regarded as an

aerial. The resultant E and H fields around it cause the cavity to oscillate.

Energy may be coupled into or from a resonant cavity by *slot* or *window* coupling. We have seen that the oscillatory H field induces currents into the walls of the cavity. These currents are always at right angles to the H field and have an amplitude proportional to the strength of the H field. When a slot is cut in the cavity such that the long side of the slot interrupts the wall currents a charge will accumulate on the long side of the slot. An oscillatory E field will therefore build up across the slot and radiation will occur in a manner similar to that of a dipole aerial but the plane of polarisation will be at right angles to the length of the slot.

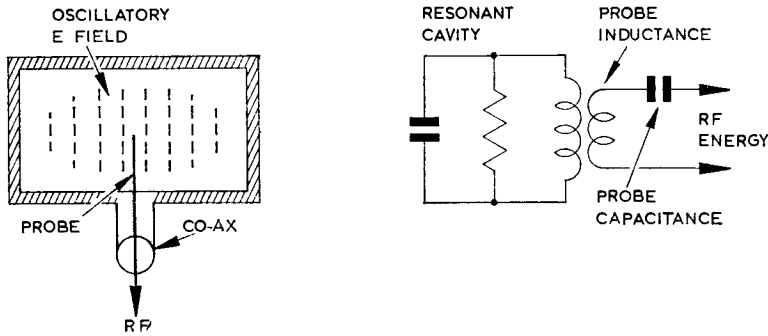


FIG. 4 PROBE COUPLING AND EQUIVALENT CIRCUIT

Fig. 5 shows slots cut at various positions in a cavity. Slots A and B interrupt the current at right angles and therefore maximum radiation occurs from these slots. Slots C and D lie parallel to the current and so radiate little energy. By feeding energy of the correct polarisation to slots such as A and B the cavity can be excited into oscillation.

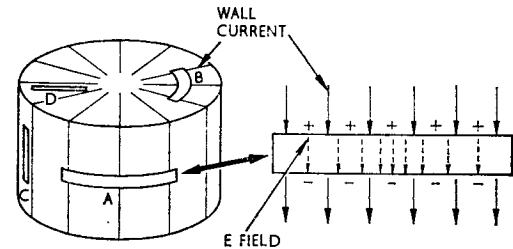


FIG. 5 SLOT COUPLING

by altering the effective inductance or capacitance of the cavity.

Inductive tuning. This is done by inserting a metal screw or *slug* into the wall of the cavity so that it distorts the magnetic field inside the cavity (Fig. 6a). By screwing the slug into the cavity the effective inductance is *reduced* and hence the resonant frequency is *increased*. More than one slug may be used to tune the cavity.

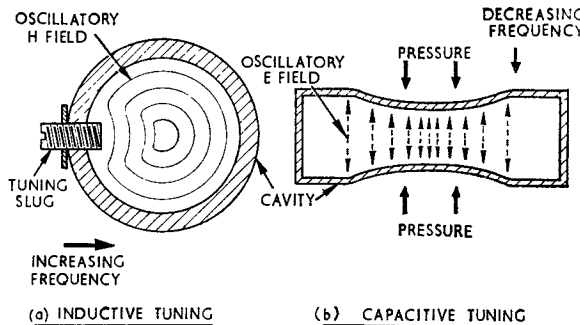
Capacitive tuning. Klystron valves operating at 3cm are usually tuned by varying the cavity *capacitance*. The roof and floor of the cavity are made flexible so that they may be moved slightly inwards, thus *increasing* the capacitance and *decreasing* the frequency slightly (Fig. 6b).

Uses of Resonant Cavities

A rhumbatron resonant cavity may be used at centimetric wavelengths as part of a klystron valve. In the

Tuning Cavity Resonators

When a cavity forms the resonant element of an amplifier or oscillator an important requirement is that it should be tunable over a small frequency range. This may be achieved



magnetron oscillator, hole-and-slot cavities may be drilled in a solid block of metal to form the resonant elements (Fig. 7). Because of their convenient size at centimetric wavelengths, resonant cavities find many other uses. We shall now consider two of the more common ones.

Absorption Wavemeter

In this device a cylindrical resonant cavity is used to measure the frequency of a centimetric input wave. The cavity may be tuned to resonance by a tuning piston which varies the internal size of the cavity. The piston is moved inside the cavity by a screw which has a calibrated micro-meter head as shown in Fig. 8. As the cavity is tuned through resonance it absorbs power from the input and a sharp dip occurs in the microammeter reading. The frequency of the input wave can then be obtained from the calibrated scale.

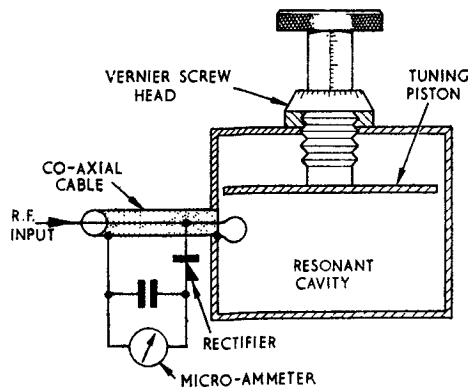


FIG. 8 CAVITY ABSORPTION WAVEMETER

gives an indication of the radar performance and shows any change in the transmitter or receiver tuning, the transmitter power output or the receiver sensitivity.

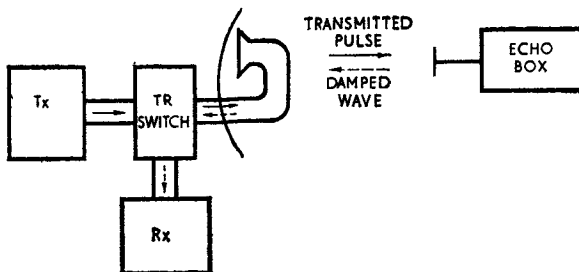


FIG. 9 USE OF THE ECHO BOX

through signal is fed to the echo box. The damped signal is then fed down the waveguide to the receiver (Fig. 10).

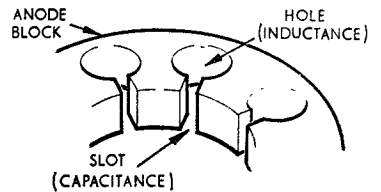


FIG. 7 HOLE-AND-SLOT MAGNETRON CAVITIES

The Echo Box

An echo box is a resonant cavity used to provide a rapid check of the overall performance of a radar transmitter and receiver. A radiated pulse from the radar transmitter is picked up by a probe or aerial and fed into an echo box tuned to the radar frequency (Fig. 9). The echo box is shock-excited into oscillation and when the transmitter pulse ends, the box re-radiates the energy as an exponentially decaying signal. This signal is picked up by the radar aerial and applied to the receiver.

The time between the end of the transmitted pulse and the instant at which the re-radiated signal falls below the minimum level required to give an indication on the radar display, is called the "ringing time". This time

The echo box test can be made as a routine check to detect any change in ringing time since the last test was made; or the overall quality of a radar set can be compared with that of a standard set by comparing the two ringing times. In these tests a standard distance between the radar aerial and the echo box aerial must be maintained. In some cases a probe is inserted into the radar receiver waveguide and a standard amount of transmitter break-

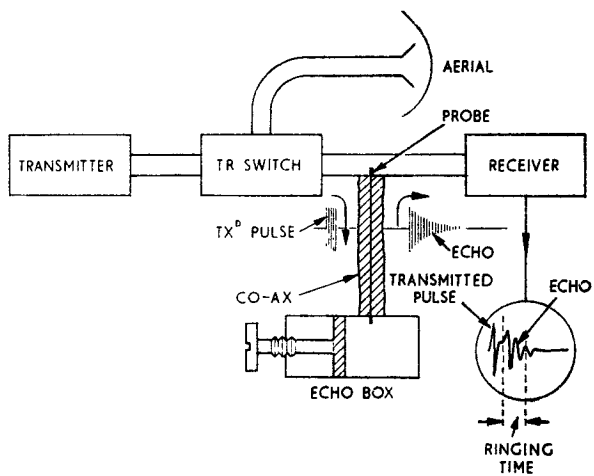


FIG. 10 ECHO BOX AND DISPLAY

CHAPTER 2

KLYSTRONS

Introduction

Klystron valves are used as microwave amplifiers and oscillators in radar transmitters and receivers. They can be designed to provide power outputs ranging from milliwatts (c.w.) to megawatts (pulsed) and in the low-power form are used as local oscillators in radar receivers; high-power klystron amplifiers are used as power amplifiers in some radar transmitters. Because they generate undesirable noise, klystrons are not used as receiver small-signal amplifiers.

The tuned circuit of the klystron amplifier or oscillator is a rhumbatron resonant cavity which may be contained wholly or partly within an evacuated glass or metal envelope. Klystrons may have one or several cavities depending on the type and purpose. Although klystrons are mainly considered as microwave devices they are also used as power amplifiers at frequencies of a few hundred megacycles per second.

Velocity Modulation

An electron in a steady d.c. electric field experiences a force which tends to accelerate it in the direction shown in Fig. 1. The electron gains kinetic energy from the d.c. field. If the field strength is increased the electron velocity is increased more and it gains more energy; if the field is reversed the electron loses energy.

In a klystron valve the cavities are usually earthed and a large negative voltage is applied to the cathode. All the electrons in the beam are therefore travelling at the same high velocity towards the hole in the first rhumbatron cavity. If the cavity is excited at its resonant frequency by an input signal the E field between the lips of the cavity will oscillate as shown in Fig. 2. In

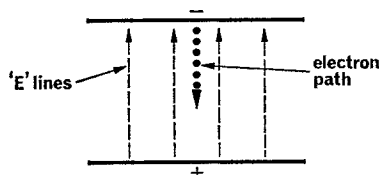


FIG 1. ELECTRON IN AN E FIELD

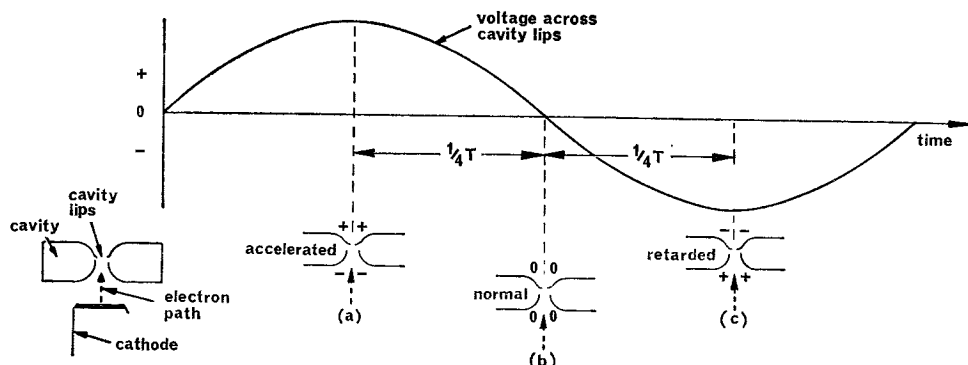


FIG 2. VELOCITY MODULATION

Fig. 2a an electron passing through the cavity gap is *accelerated* and it gains energy from the oscillatory field. A quarter of a cycle later (Fig. 2b) there is no field across the gap and the electron emerges with unchanged velocity. A further quarter of a cycle later (Fig. 2c) an electron will be *retarded* and some of its kinetic energy will be given to the oscillatory field.

Having passed through the cavity gap the electrons in the stream are travelling at different velocities, i.e. they have been *velocity modulated*. The faster-moving electrons will at some point catch up with the slower-moving ones which passed through the cavity gap before them and a

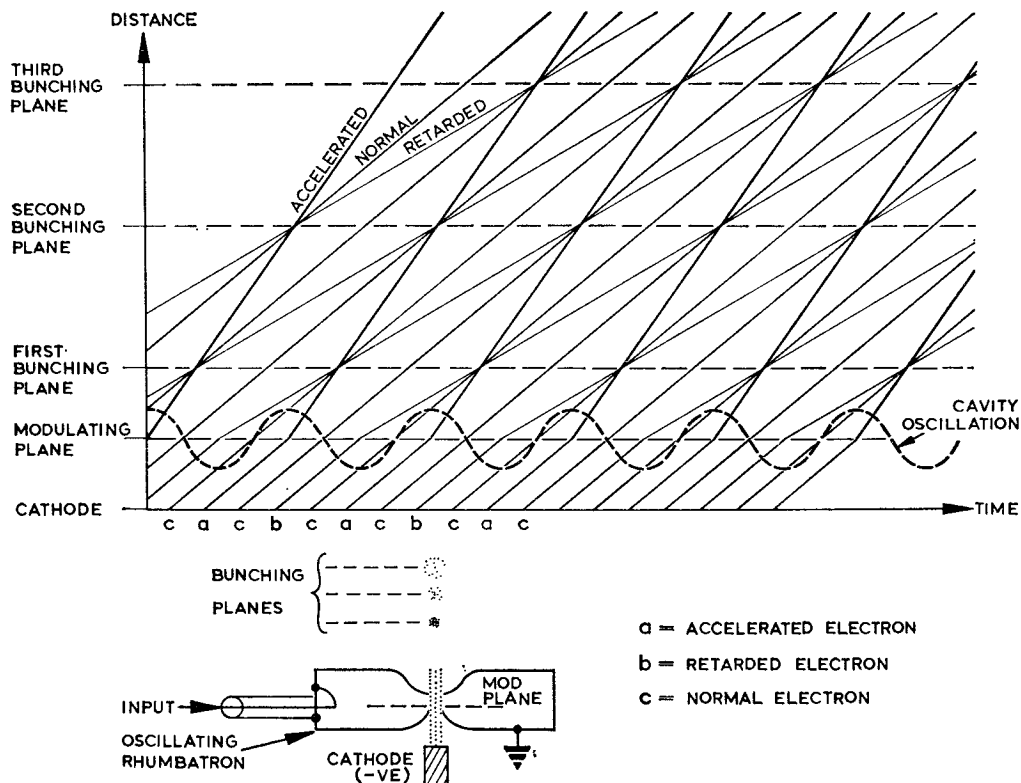


FIG 3. THE APPLGATE DIAGRAM

bunch of electrons will be formed at that distance from the modulating plane. This is called a *bunching plane*. As the distance from the modulating plane increases another bunch is formed, i.e. there are several bunching planes. The Applegate diagram (Fig. 3) is a plot of distance against time and illustrates the action. It can be seen that at each plane a bunch is formed once per cycle of cavity oscillation.

If the voltage between the cathode and the cavity is decreased, the bunching planes occur nearer the cavity and bunches are formed later in time. Owing to mutual repulsion between electrons, the higher bunching planes are less well defined, i.e. the bunches of electrons are much looser and larger.

The Double Cavity Klystron

The principle of velocity modulation described in the previous paragraphs is used in the *double cavity* klystron, the construction of which is shown in Fig. 4. The indirectly heated cathode is held at a voltage negative with respect to the cavities which are usually earthed.

After passing through the cavities the electrons are returned to the cathode via the collector electrode and the power supplies. The cavity near the cathode is called the *buncher* cavity and the second cavity is the *catcher*. The space between the two, where the bunches of electrons form, is the *drift space*.

The buncher cavity is excited by the input signal and causes velocity modulation of the electron beam. The catcher cavity is placed at one of the bunching planes and is tuned to the same frequency as the buncher cavity. If a maximum retarding field occurs when a bunch of electrons arrives at the catcher cavity gap, maximum energy is extracted from the electron beam.

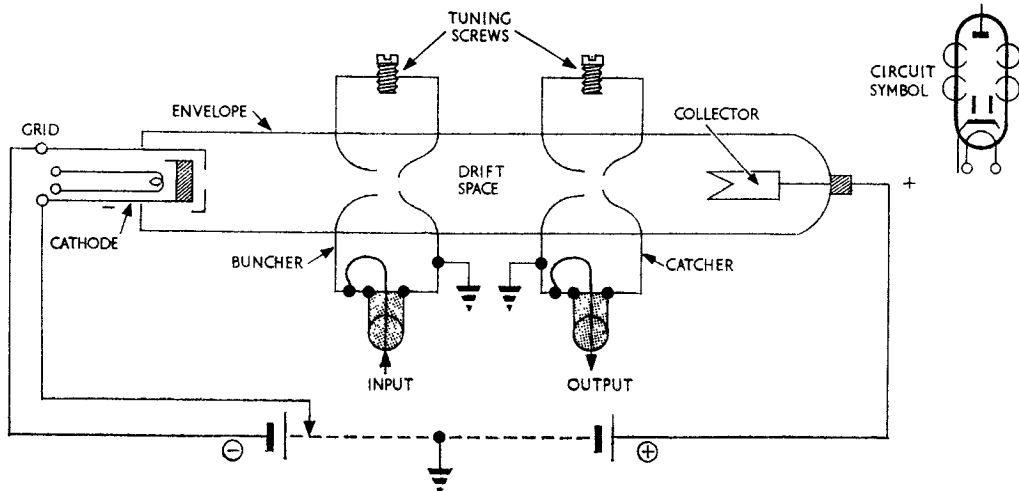


FIG. 4 DOUBLE CAVITY KLYSTRON

Half a cycle later when the catcher cavity presents an accelerating E field, there is no bunch at the gap, few electrons are accelerated and little energy is extracted from the cavity field. Therefore, over one cycle of oscillation more electrons have given energy to the catcher cavity than have taken energy from it; thus the electron beam has given energy to the catcher cavity in the correct phase to enable oscillations to build up. These conditions can be obtained by placing the catcher cavity at the correct distance from the buncher cavity and by adjusting the klystron d.c. voltage and the amplitude of input signal.

An amplified output may be taken via a loop in the catcher cavity and applied to the load. If the input signal is removed some of the amplified energy in the catcher cavity can be fed back into the buncher cavity by loop couplers. In this way the double cavity klystron will operate as an oscillator.

Multi-cavity klystrons. High-power klystron amplifiers are sometimes used in c.w. or pulsed radar transmitters and in microwave links. These klystrons have an additional cavity inserted between the buncher and catcher cavities. The second cavity is placed at the optimum drift distance and since it is unloaded it has a high Q and develops large voltages across its gap. Thus the beam is velocity modulated again and increased power is developed in the catcher cavity. Four or five cavities may be added in this way. The cavities may be stagger-tuned (as the tuned circuits in an i.f. amplifier) giving an increased bandwidth. A multi-cavity 3 cm klystron amplifier which gives a c.w. output of 20kW is shown in Fig. 5.

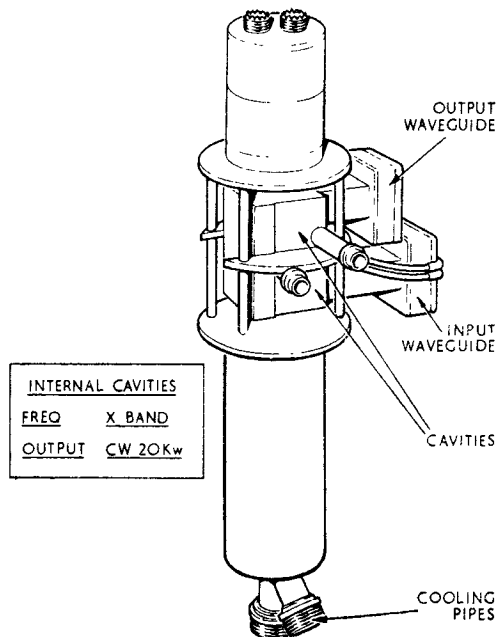


FIG. 5 MULTI-CAVITY CW KLYSTRON

(A.L. 2, August, 1965)

The kinetic energy in the electron beam is converted into heat when the electrons strike the collector and some high-power klystron amplifiers employ a water-cooled

collector. This heat represents wasted d.c. energy and to improve efficiency the collector voltage may be decreased, so reducing the kinetic energy of the electrons at the collector.

The physical size of klystrons depends on the size of the cavities and hence upon the wavelength. Thus, low-frequency klystrons (100—1,000 Mc/s) have large cavities with large gaps, enabling high-power electron beams to be used.

Cavities may have metal "grids" or meshes across the cavity gaps to strengthen the oscillatory E field. This results in increased efficiency, but as the grids intercept some of the electrons, the noise level is increased.

Klystron amplifiers may be driven by a c.w. oscillator and the power supplies pulsed to give a more easily controlled pulse shape than can be obtained with a magnetron. Also, the frequency can be controlled, over a narrow band, better than in a magnetron.

Multi-cavity klystrons may have a power gain of 40 db. over a bandwidth of a few megacycles per second with an efficiency of about 40%. Output power levels for high-power klystrons operating in the S band (10cms) are typically a few kilowatts c.w. or several megawatts pulsed.

The size of klystrons ranges from about two inches for X band (3cm) receiver local oscillator klystrons to several feet for high-power multi-cavity transmitter klystrons working in the L band (25cm). In some high-power klystrons the electrons strike the collector at such high speeds that X-rays are emitted and these klystrons have to be shielded.

The Reflex Klystron

The double cavity klystron amplifier can be made into an oscillator by feeding some energy from the output cavity back into the input. The *reflex klystron* is a simpler type of velocity modulation oscillator in that one cavity serves as both buncher and catcher and there is no need for an external feedback loop. The construction of a 10cm reflex klystron is shown in Fig. 6.

As in the klystron amplifier the cathode is taken to a negative voltage and the cavity is earthed. Thus electrons are accelerated from the cathode towards the cavity and shock-excite the cavity into oscillation. The electron beam is then velocity modulated by the cavity oscillatory E field. The electrons pass through the cavity gap and come under the influence of a repelling d.c. E field due to the *negative* voltage, relative to the cavity, on the *repeller* electrode. This voltage is sufficiently high to cause the electrons to stop and turn back towards the cavity. Thus the electrons *return* through the cavity gap.

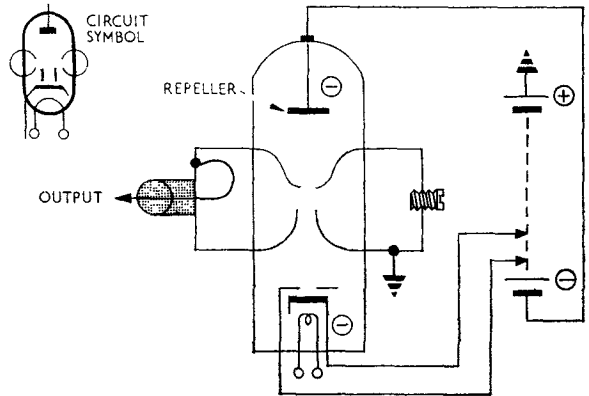


FIG. 6 THE REFLEX KLYSTRON

The distance travelled by an electron before being turned back to the cavity depends on the repeller voltage and the velocity of the electrons. Thus electrons which are accelerated by the oscillatory E field will move nearer to the repeller than retarded electrons and may arrive back at the cavity gap at the same time as retarded electrons which were emitted later. Bunches of electrons will be formed at the cavity gap on the backward path if the distance of the repeller from the gap and the repeller voltage are correctly chosen.

In the modified Applegate diagram of Fig. 7 the repeller voltage has been adjusted so that the transit times of electrons *a*, *b* and *c* are equal to 1, $\frac{3}{4}$ and $\frac{1}{2}$ a period of cavity oscillation respectively. These three electrons form a bunch at the cavity gap on the downward path when the

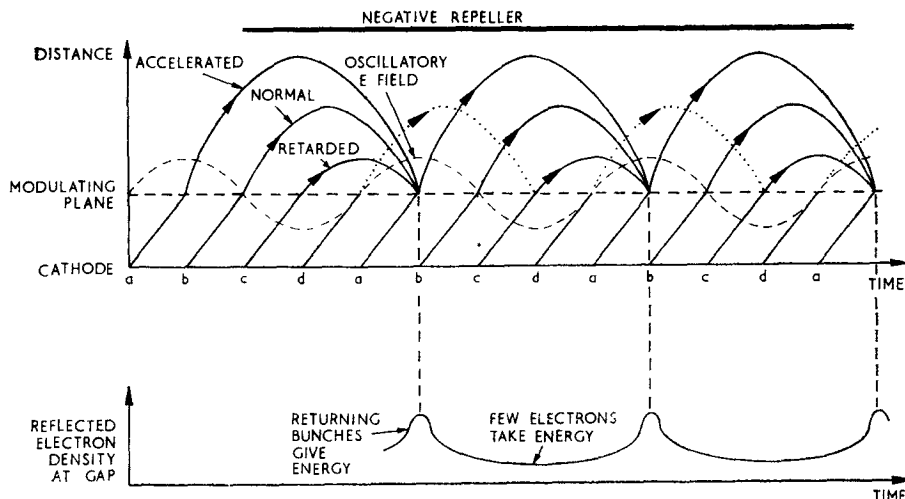


FIG. 7 REFLEX KLYSTRON APPLEGATE DIAGRAM

cavity is presenting a retarding E field to *backward* travelling electrons. Thus some of the energy possessed by these three electrons is given to the cavity. Electron *d* arrives back at the cavity gap and is accelerated by the E field and thus takes energy from the cavity. However, over one complete cycle, more energy is given to the cavity than is taken from it and oscillations are sustained.

Voltage modes of oscillation. Notice that in Fig. 7 the bunches of electrons arrive back at the cavity gap three-quarters of a period of oscillation after the "normal" electron (electron *b*) has passed upward through the gap. The klystron is said to be operating in the *first voltage mode*.

If the repeller voltage is made less negative than indicated in Fig. 7 the electrons will travel further towards the repeller. By suitable adjustment of the repeller voltage the bunches can be made to form at the cavity gap $1\frac{1}{2}$, $2\frac{1}{2}$, $3\frac{1}{2}$, etc. of a period after the "normal" electron in each

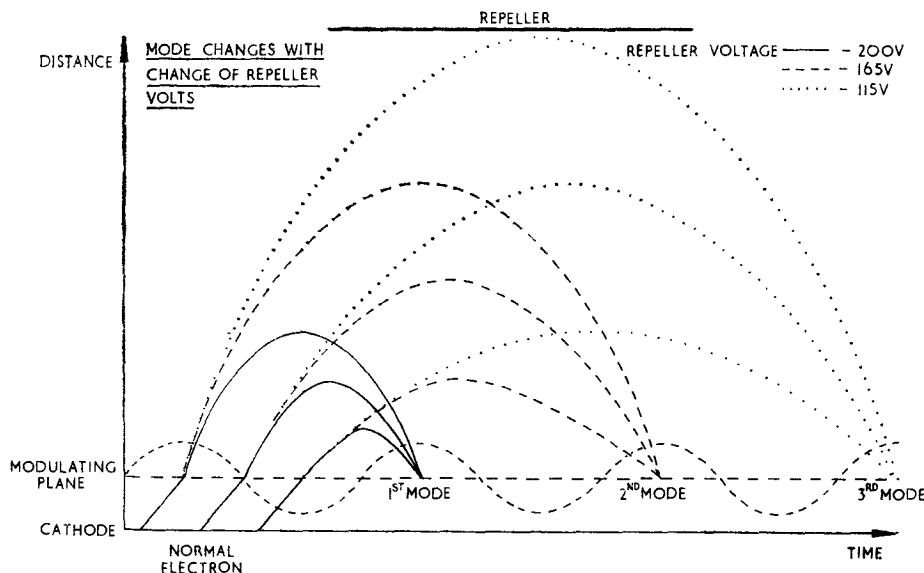


FIG. 8 REFLEX KLYSTRON VOLTAGE MODES

bunch has passed upward through the cavity gap. For each setting of the repeller voltage, oscillations will be sustained and the klystron is said to be oscillating in the second, third, fourth, etc. voltage mode (Fig. 8).

Power output and frequency. In the higher voltage modes the electron bunches are less well defined and they give less energy to the cavity. Thus the power output of the klystron is reduced. Fig. 9a shows how the power output falls in the higher voltage modes. Transit time considerations generally prohibit the use of the first mode.

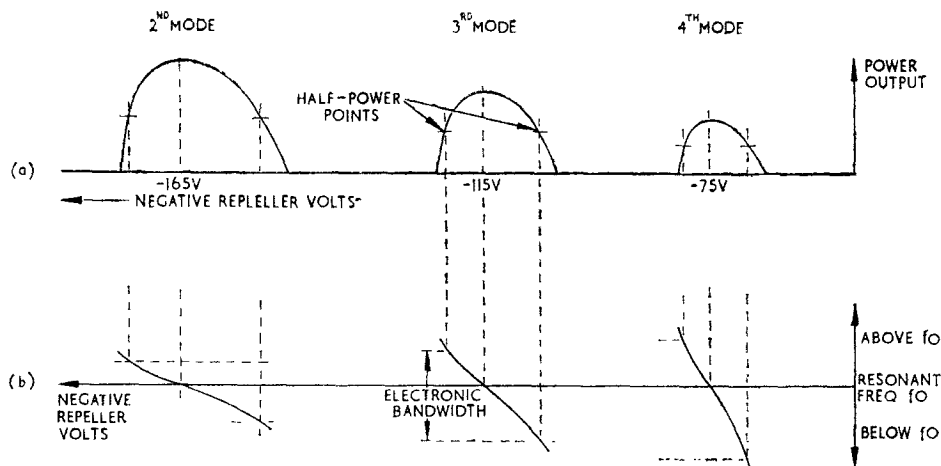


FIG. 9 CHANGE OF POWER OUTPUT AND FREQUENCY WITH CHANGE OF REPELLER VOLTS

The resonant frequency of a reflex klystron may be varied by mechanically altering the shape or size of the cavity with capacitive or inductive tuning screws. It is also possible to vary the output frequency by varying the repeller voltage slightly about the optimum value for the mode. If the repeller voltage is made slightly *more negative*, electron bunches form earlier and the output frequency *increases*. A *less negative* repeller voltage results in a *decrease* in frequency (Fig. 9b). As the frequency moves from resonance the power output falls. The change of frequency caused by a change of repeller voltage, between the half-power points, is called the *electronic bandwidth* of the klystron.

Fig. 9 shows that by operating the klystron in a higher mode a wider bandwidth is obtained, although the power output is reduced. A typical electronic bandwidth for a 3cm reflex klystron is 30Mc/s. When a reflex klystron is used as the local oscillator in a radar superhet receiver, a power output of less than 1mW is required. However, in order to maintain the i.f. constant it may be necessary to vary the klystron frequency as the radar transmitter frequency varies, and so a wide klystron electronic bandwidth is desirable. By operating the klystron in one of the higher modes this requirement is met. A reflex klystron can also be used to provide a frequency modulated output by applying the modulating voltage to the repeller electrode.

The reflex klystron is a low-power oscillator with an efficiency of only a few per cent. It is widely used as a centimetric local oscillator in radar receivers and may have an internal cavity or an external cavity. The electrodes are brought to a conventional valve base (octal, B7G, etc.) and the output may be fed directly into a waveguide or along coaxial cable. Fig. 10a shows a 10cm and Fig. 10b a 3cm reflex klystron. Characteristics are given in Table 1.

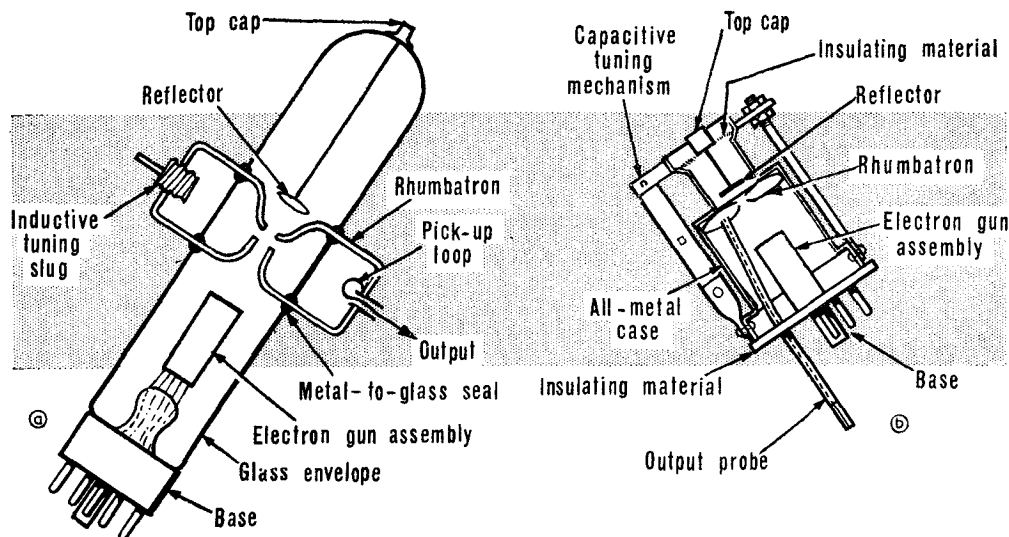


FIG. 10 REFLEX KLYSTRONS

Cavity Type	Base	Freq Range (Gc/s)	Accele- rating Volts	Beam Current (mA)	Repeller Volts	Power Output	Elect. Bandwidth	Output
External	B7G	8.2-11.7	350	40	-350	130mW	20Mc/s	Waveguide
External	Octal	3.2-3.4	250	32	-140	150mW	30Mc/s	Co-Ax
Internal	B8G	4.4-4.8	750	143	-290	3.7W	50Mc/s	Co-Ax

TABLE 1. DETAILS OF REFLEX KLYSTRONS

CHAPTER 3

TRAVELLING-WAVE TUBES

Introduction

To meet the requirements of different radar systems microwave amplifiers and oscillators other than multi-cavity and reflex klystrons have been developed. In this chapter we shall consider the properties and applications of two of these devices.

The Travelling-wave Tube Amplifier

An essential part of a klystron amplifier or oscillator valve is the resonant cavity which normally has a very high Q-factor. The high-Q cavity is necessary in order to apply a magnified oscillatory voltage across a very short gap to modulate the electron velocity. The use of high-Q cavities means that the bandwidth of the klystron is restricted and in many applications this is a considerable disadvantage.

In the reflex and double-cavity klystrons an electron has only one chance of transferring its energy to the cavity. This accounts for the fairly low efficiency of these devices. In the multi-cavity klystron the electron gives energy to each cavity it passes through and this results in an increased efficiency.

In the travelling-wave tube (t.w.t.) an electron beam is velocity modulated but the modulation is a continuous process in which electrons give energy over many periods of the input signal. This results in a high efficiency. The t.w.t. does not employ resonant cavities and is a broad-band device. Perhaps the most important feature of the t.w.t. is its *low noise factor* which is much lower than that of the klystron. These properties make the t.w.t. a suitable receiver and transmitter microwave amplifier.

The constructional details of a t.w.t. amplifier are shown in Fig. 1. An electron gun emits

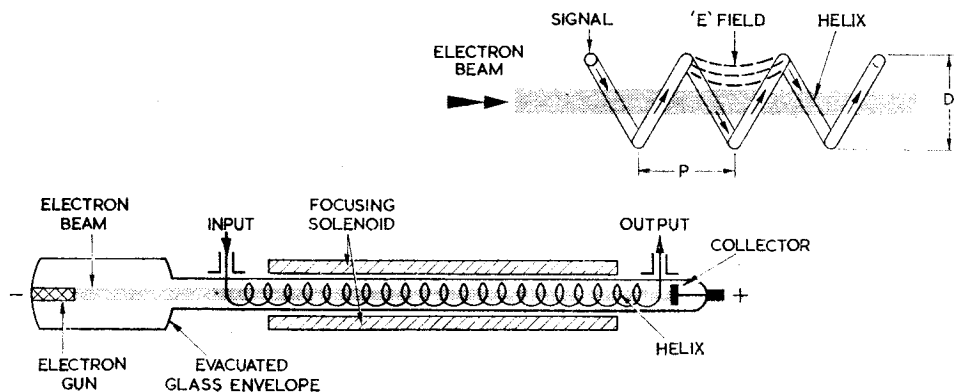


FIG. 1 DETAILS OF THE TRAVELLING-WAVE TUBE

electrons which are attracted through the tube to a positive collector. The input signal is applied to one end of a wire, wound in the form of a *helix*, the amplified output being taken from the other end of the helix. The electrons travel down the axis of the helix and to prevent their striking it, an axial magnetic field is provided by a *focusing solenoid*. The helix slows down the speed of the wave along the axis of the tube and is therefore called a *slow wave structure*.

Operation. The small-amplitude input signal applied to the input coupler travels along the helix wire with a velocity almost equal to that of light. This travelling-wave produces an E field along the axis of the helix. The *axial* velocity of this field is much less than that of light (c), being equal to $\frac{c \times \text{pitch of helix}}{\text{circumference}}$: the design of the tube is such that the velocity of the electron

beam is approximately the same as that of the axial E field set up by the signal on the helix. Thus an electron at the input end of the helix that experiences an accelerating field will move with this field and experience an accelerating force over many periods of the alternating signal. Half a cycle later an electron at the beginning of the helix will encounter a decelerating field and will be decelerated over many periods.

Thus the electron beam is velocity modulated and bunches of electrons are formed as in the klystron. These bunches induce a voltage in the helix which augments the signal voltage. The amplified signal causes more bunching which further increases the amplitude of the signal on the helix. In this way the d.c. energy in the electron beam is converted into a.c. energy in the travelling-wave signal and the tube acts as an amplifier.

Imperfections in the input and output couplers may cause reflections along the helix which are amplified in the forward direction. If the amplitude of the reflections and the forward gain of the tube are sufficient, instability results and the tube may act as an oscillator. To prevent this the inside of the glass tube near the centre may be coated with graphite (Fig. 2). The graphite acts as an attenuator both to the reflected and to the forward travelling-waves but as it does not affect the bunched electrons the forward wave continues to be amplified while the reflected wave dies away.

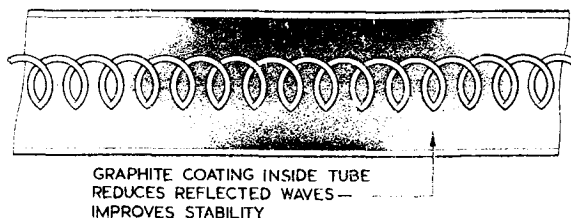


FIG. 2 TRAVELLING-WAVE TUBE ATTENUATOR

The longer the helix the more interaction there is between electrons and travelling-wave and so the greater the gain of the tube. However, as energy is extracted from the electrons so the mean velocity of the beam is decreased and if the helix is made too long the velocity of the travelling-wave may become greater than that of the electron beam towards the end of the helix. Synchronism between the wave and electrons is lost and electron bunches may *extract* energy from the wave with a resultant fall in gain. For this reason the velocity of the electron beam at the cathode end of the helix is made slightly *greater* than the axial E field velocity. Thus reasonable synchronism is obtained along most of the helix.

The longitudinal d.c. magnetic field produced by the focusing solenoid is necessary in order to keep the electron beam at the centre of the helix throughout the length of the tube. It thus prevents the electrons from striking the wire helix and causing wastage in power, heating in the helix and considerable noise in the output.

Solenoids and the necessary power supplies are bulky, heavy and expensive. A considerable saving in weight is achieved by replacing the solenoid with small permanent magnets placed at intervals along the tube. The beam is compressed over a short length of tube and then expands due to mutual repulsion between electrons. Once again it is compressed, then expands and so the sequence continues (Fig. 3). This is known as *periodic focusing* and may be achieved electrostatically as well as magnetically.

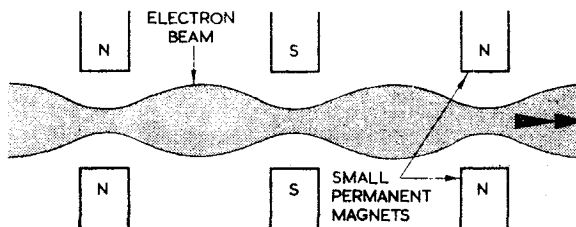


FIG. 3 PERMANENT MAGNET PERIODIC FOCUSING

It should be noted that the axial magnetic field plays no part in the conversion of energy from d.c. to a.c.; its only function is to focus the beam.

Low-power t.w.t.'s have been designed with noise factors as low as 2 to 3 db and this makes them suitable wide-band signal amplifiers in microwave receivers. A typical low-power t.w.t. is

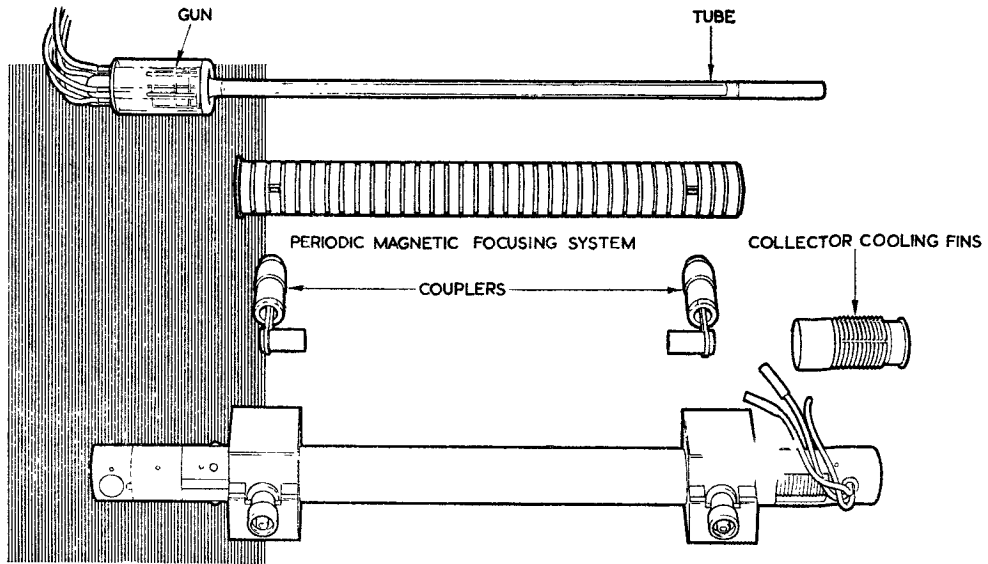


FIG. 4 A LOW-POWER TRAVELLING-WAVE TUBE

shown in Fig. 4. Periodic magnetic focusing is used and cooling fins are employed to dissipate the heat generated at the collector. Energy is coupled into and out of the tube by coaxial line; some tubes use waveguide coupling.

High-power t.w.t.'s with power gains of up to 40db and power outputs of 1kW c.w. or 2MW pulsed are used in radar transmitters. They generally employ a more robust form of slow wave structure than the wire helix and the bandwidth is somewhat narrower than low-power tubes; they may be amplitude or frequency modulated. Details of low- and high-power t.w.t.'s are given in Table 1.

Use	Freq Range (Gc/s)	Collector Voltage (V)	Collector current (mA)	Power Output	Power Gain (db)
Broadband amplifier ..	7-12	1,200	0.6	6 mW	25
Low noise amplifier ..	3-4	500	0.2	3mW	25
Power Amplifier ..	7-8	1,800	40	10W	35

TABLE 1. DETAILS OF TRAVELLING-WAVE TUBE AMPLIFIERS.

The O-type Backward-wave Oscillator

The *O-type backward-wave oscillator* (b.w.o.) or *O-carcinotron* is a development of the t.w.t. amplifier. An electron beam delivers energy to a wave travelling *backwards* along a slow wave structure, i.e. the wave travels in the opposite direction to the electrons. The amplified backward wave velocity-modulates the electron beam and also provides the output energy of the oscillator. The design of the tube is such that the forward travelling-wave is not amplified.

One type of slow wave structure often employed in the b.w.o. is the folded transmission line

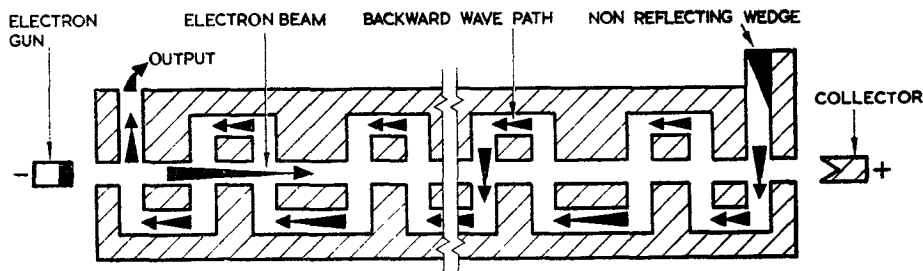


FIG. 5 O-TYPE BACKWARD WAVE OSCILLATOR USING INTER DIGITAL LINE

or *interdigital line* shown in Fig. 5. This consists of two metal combs, bolted together so that the teeth interlace; the combs are drilled axially to provide a path for the electron beam. The electron gun emits electrons which are attracted through the slow wave structure towards a positive collector.

As electrons pass a gap in the transmission line, an oscillatory E field is set up across the gap and a backward travelling wave moves down the slow wave structure, tacking from side to side across the electron stream. A forward wave moving towards the collector is absorbed by the non-reflecting termination and plays no further part in the action. Interaction between the electron beam and the backward wave causes velocity modulation of the electrons and the bunches formed give energy to the wave.

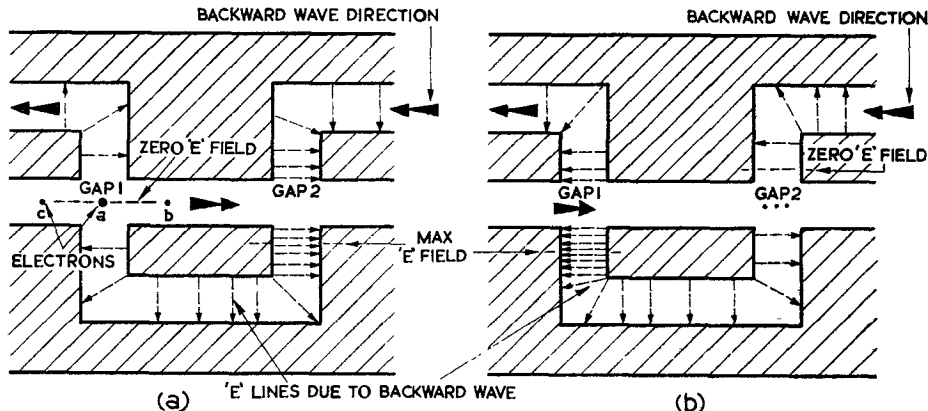


FIG. 6 VELOCITY MODULATION AND ENERGY TRANSFER IN AN O-BWO

Fig. 6a shows a section of the slow wave structure and the field pattern set up by a backward wave at one instant of time. The field at gap 1 is zero and about to become accelerating. Thus the velocity of electron *a* is unaffected, electron *b* was decelerated when it passed the gap earlier and electron *c* will be accelerated. The electrons are thus velocity modulated and will form into bunches. The velocity of the electron bunches is controlled by the collector d.c. voltage and this is adjusted such that the time taken for the electrons to move from gap 1 to gap 2 is slightly less than the time taken for the wave to travel from gap 2 to gap 1. A bunch of electrons forms at gap 2 and encounters a slightly decelerating field as shown in Fig 6b. Energy is transferred from the electron bunch to the travelling wave and further bunching takes place.

As the wave gets progressively nearer the output so the retarding E field presented to the bunches at the gaps intensifies and the wave amplitude increases. Also, as the electrons approach the collector, bunching increases and this results in more energy being given to a somewhat weaker field. In this way energy is transferred from electron bunches to the travelling wave at each gap. The degree of bunching and the amplitude of the backward wave vary with distance along the tube as shown in Fig. 7.

In the b.w.o. the amplified backward travelling-wave near the cathode acts as the signal in a normal t.w.t. amplifier and velocity modulates the beam. Thus the b.w.o. is a type of t.w.t. amplifier with a special form of "built in" feedback, i.e. it is an oscillator.

The frequency at which the b.w.o. oscillates depends on the transit time of the electrons between gaps in the slow wave structure, i.e. on the voltage applied to the collector. By varying the collector voltage the output frequency can be varied over a wide range. This is the big advantage of the b.w.o.; it is possible to obtain a 2:1 frequency range with a voltage variation of about 10:1.

The b.w.o. is suitable as a low-power local oscillator in a wide-band microwave receiver. Power outputs of a few milliwatts are possible at the higher frequencies. Relatively high powers, of the order of watts, can be obtained at S band frequencies (3Gc/s). Efficiencies are about the same as for t.w.t. amplifiers. Other forms of slow wave structure can be used including a special form of helix.

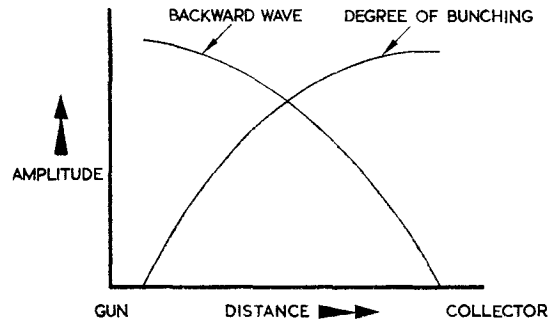


FIG. 7 AMPLITUDE OF BACKWARD WAVE AND ELECTRON DISTRIBUTION IN O-BWO

CHAPTER 4

MAGNETRONS

Introduction

A magnetron is a high-power microwave oscillator which forms the basis of most centimetric radar transmitters. Oscillations are maintained by the transfer, via electrons, of d.c. energy to an oscillatory field. As in the klystron and t.w.t., electrons gain energy by being accelerated by a strong d.c. E field and give some of this energy to a retarding oscillatory E field. In the magnetron, however, the path of the electrons is influenced by a strong d.c. magnetic field produced by a permanent magnet. This type of interaction, where a d.c. H field plays a part in the transfer of energy, is called M-type (magnetic) interaction as distinct from the klystron and t.w.t. O-type (ordinary) interaction.

Effect of H field

An electron moving in a steady d.c. magnetic field as shown in Fig. 1a will have a force exerted on it. The magnitude of this force depends on the electron speed and the strength of the

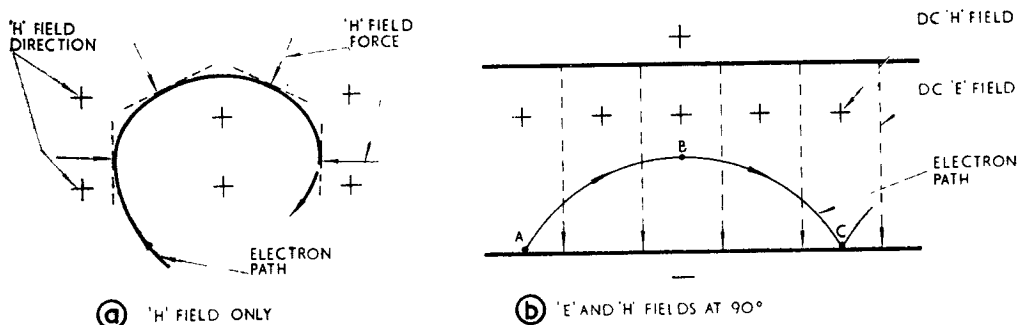


FIG. 1 ELECTRON PATHS

H field; the direction of the force is always at 90° to the electron path and is given by Fleming's left hand rule. The electron thus moves in a circular path (Fig. 1a).

In the magnetron, electrons move under the influence of a steady d.c. E field and a steady d.c. H field, these fields being at right angles to each other. Fig. 1b shows the path taken by an electron influenced by these two fields. A stationary electron at A is not influenced by the H field but is accelerated towards the positive plate by the E field. As soon as it starts to move the H field exerts a force on it at right angles to its direction and it moves in a curved path to B. From A to B the electron is accelerated by the E field but from B to C it is retarded. At C, it momentarily comes to rest and is then accelerated towards the positive plate and continues its curved path. The path ABC taken by the electron is a *cycloid*; the curvature of the cycloid depends on the strength of the E and H fields.

When an H field lies parallel to the cathode of a cylindrical-anode diode as shown in Fig 2a the normal radial path (AB) of the electrons becomes curved (CD) owing to the forces discussed in the previous paragraph. As the strength of the H field is increased the curvature of the cycloid is increased. When a critical value of H field is reached the electrons fail to reach the anode and are returned to the cathode (EF); anode current ceases. This critical value of H field depends on the voltage between anode and cathode (Fig. 2b).

If the velocity of an electron is increased at, say, point x (Fig. 2a) the force at right angles to its path will increase causing the electron to move in the direction y. If the electron is retarded the curvature of its path decreases (path z).

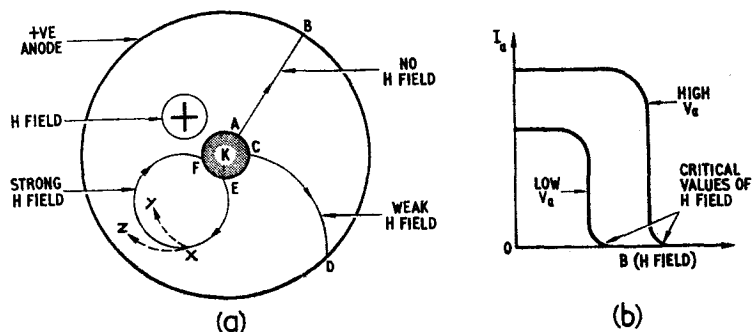


FIG. 2 EFFECT OF H FIELD ON AN ELECTRON IN A DIODE

The Split-anode Magnetron

In order to understand the principles of a practical multi-cavity magnetron we shall first consider the action of the simple device shown in Fig. 3. This consists of a diode valve with a cylindrical anode split into two segments; a LC circuit is connected between the two segments. A strong steady d.c. H field parallel to the cathode is provided by a permanent magnet. When the h.t. is switched on, the tuned circuit oscillates and an oscillatory E field is set up between the anode segments. The steady d.c. E and H fields are not shown in Fig. 3, although they are present.

Electrons emitted from the cathode move towards the anode in a curved path. Let us consider electron A (Fig. 3) which is accelerated towards the anode by the steady d.c. E field, gaining energy from this field. Its cycloid path takes it through the oscillatory E field at X and if this field is a retarding field the electron gives energy to the tuned circuit. Its velocity decreases and the curvature of its path decreases until it comes to rest at B.

It is again accelerated towards the anode by the steady d.c. E field and enters the oscillatory E field at Y. If the transit time of the electron is equal to half a period of E field oscillation the field at Y is now a retarding field and the electron again gives energy to the tuned circuit before it comes to rest at C. This action occurs several times before the electron finally reaches the anode.

Electron D, emitted at the same instant as electron A, also gains energy as it is accelerated towards the anode but it enters the oscillatory E field at Y when this field is an accelerating field. Thus it extracts energy from the tuned circuit, the curvature of its path increases and it returns to the cathode where it gives up its energy in the form of heat.

Since electrons are emitted evenly from the cathode surface there will be approximately as many "D-type" electrons as there are "A-type". However, since the A-type electrons perform several cycloids they give more energy to the tuned circuit than is extracted from it by the one-cycloid D-type electrons. Thus oscillations are maintained in the tuned circuit.

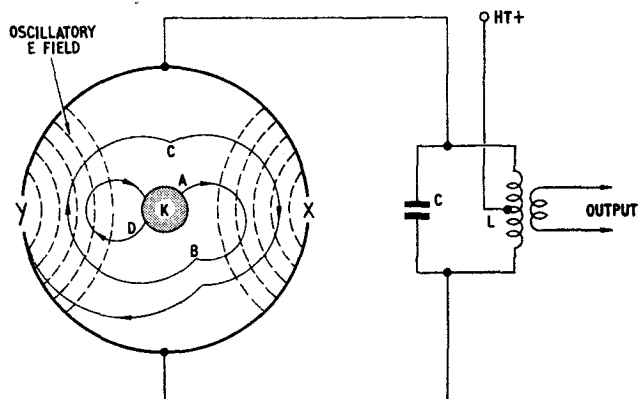


FIG. 3 THE SPLIT-ANODE MAGNETRON

The Multi-cavity Magnetron

The efficiency of the simple split-anode magnetron is quite low but it can be improved by increasing the number of retarding E fields through which the electrons can pass, i.e. by increasing the number of anode segments. This is effectively done in the *multi-cavity* magnetron oscillator, a cross-section of which is shown in Fig. 4.

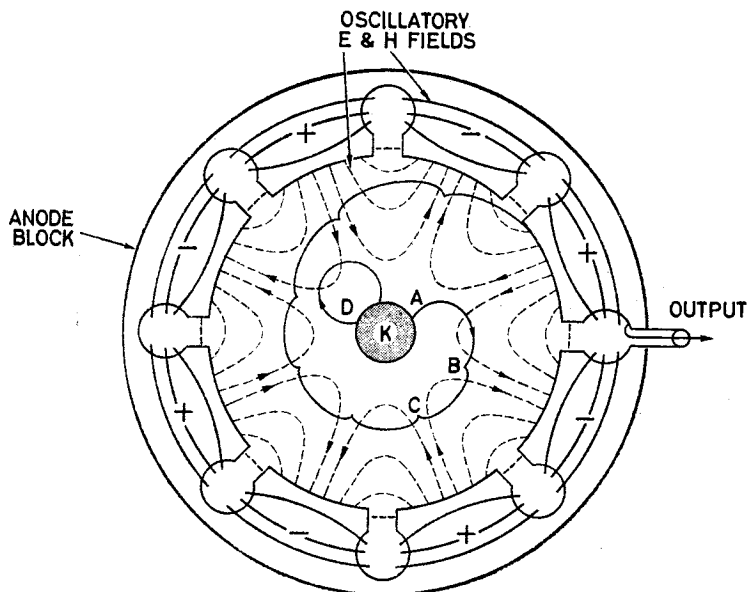


FIG. 4 MULTI-CAVITY MAGNETRON

It consists of a hollow cylindrical copper block into which is drilled a number of cavities (usually 8, 12 or 16). Each cavity resonates at the same frequency, determined by its dimensions. The cathode is mounted in the centre of the block and being oxide-coated is capable of heavy emission. Top and bottom plates seal the magnetron and the whole structure is evacuated. A strong steady d.c. E field is obtained by applying a negative e.h.t. voltage to the cathode while the anode is held at earth. A steady d.c. magnetic field parallel to the cathode is provided by a strong permanent magnet. In Fig. 4 only the oscillatory E and H fields are shown.

When the magnetron is oscillating the cavities are coupled together by oscillatory E and H fields and thus energy is extracted from all cavities by a loop or slot cut in one cavity. A typical oscillatory field pattern inside an oscillating magnetron is shown in Fig. 4.

An electron emitted from the cathode at A will be retarded by the oscillatory E field during its cycloid path AB. The transit time is made equal to half a period of magnetron oscillation so that the direction of all oscillatory fields will be reversed and the electron again retarded during cycloid BC. Several further cycloid paths are formed before the electron reaches the anode and during each cycloid, energy is transferred by the electron from the d.c. source to the oscillatory field.

An electron emitted at D (Fig. 4) at the same instant as that from A will be accelerated by the oscillatory E field and returns, after one cycloid hop, to the cathode. The action of electrons A and D is similar to that in the split-anode magnetron but as resonant cavities are used in conjunction with a steady e.h.t. voltage (about 10kV) and a strong steady magnetic field, oscillations are in the microwave band. Also, since the A-type electrons encounter more oscillatory

E fields, they transfer more energy from the e.h.t. supply to the oscillating cavities, and a very high power output with high efficiency is obtained.

In addition to this basic action, the electrons in a multi-cavity magnetron are velocity modulated and tend to form bunches as they pass the cavities. In an 8 cavity magnetron these bunches form four "spokes" centred on the cathode and rotating in synchronism with the oscillatory field. Thus clouds of electrons pass the cavity gaps at the instant of maximum retarding oscillatory field and a large amount of energy is given to the cavity.

A breakaway view of a 3cm multi-cavity magnetron oscillator is shown in Fig. 5. The permanent magnet is shown separately. Cooling fins are necessary in order to dissipate the heat caused by electrons striking the anode. Magnetrons vary in appearance and size depending on the wavelength and power output. The output coupling may feed into waveguide or, at longer

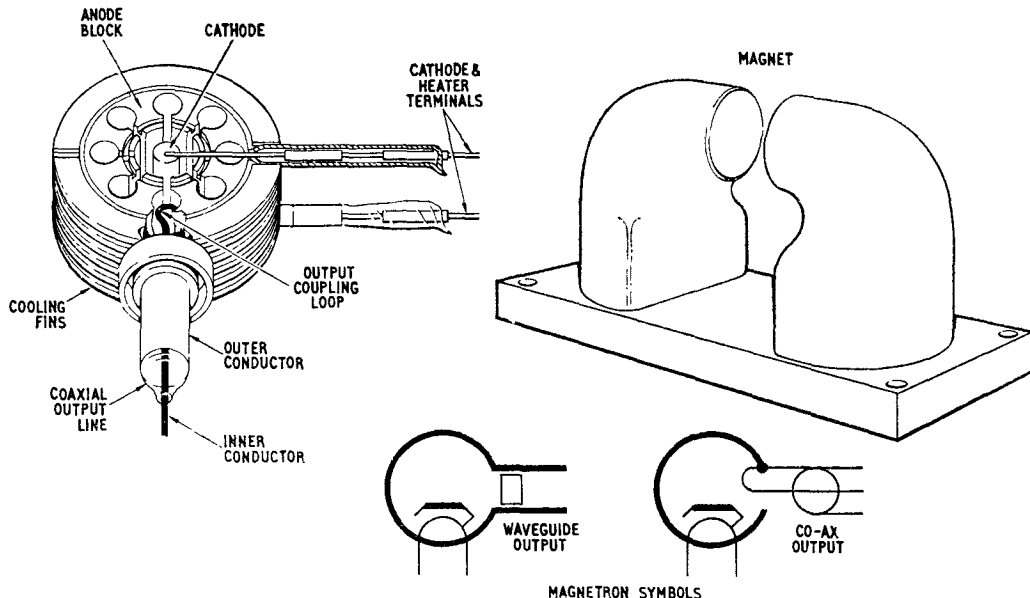


FIG. 5 3CM MULTI-CAVITY MAGNETRON AND SYMBOLS

wavelength, into co-axial cable. Some "package-type" magnetrons are built with the magnet and anode block as a single unit.

Electrons which are returned to the cathode after one cycloid give up their kinetic energy as heat and in some magnetrons the heater current must be reduced or cut off, once the magnetron is oscillating, to prevent overheating of the cathode.

Magnetron Strapping

Oscillations are started in a multi-cavity magnetron by orbiting electrons which induce a charge into the cavities and this shock-excites them into oscillation. The oscillatory field pattern shown in Fig. 4 is not the only one that can be set up in a magnetron but it is the most efficient pattern since the orbiting electrons pass through the largest possible number of oscillatory E fields.

These field patterns are called *modes* and since adjacent segments of the magnetron in Fig. 4 are 180° or π radians out of phase this magnetron is said to be operating in the " π " mode. Other modes in which the magnetron may oscillate are shown in Fig. 6. When operating in one of these other modes the power output of the magnetron falls and the frequency changes from that of the π mode.

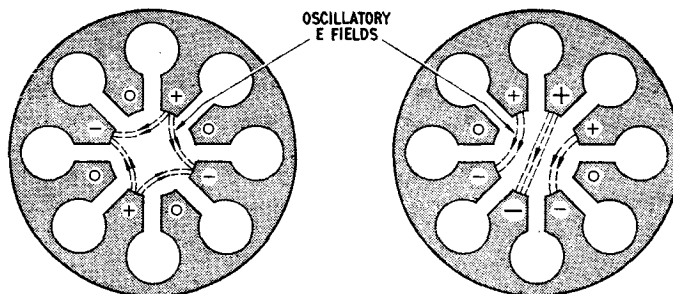


FIG. 6 MAGNETRON MODES OF OSCILLATION

To encourage the magnetron to operate in the most efficient mode, i.e. the π mode, alternate segments of the anode block are strapped together with copper wire. Two methods of strapping are shown in Fig 7a and b. At wavelengths shorter than 3cm the "rising sun" form of anode block cavities is used to suppress unwanted modes.

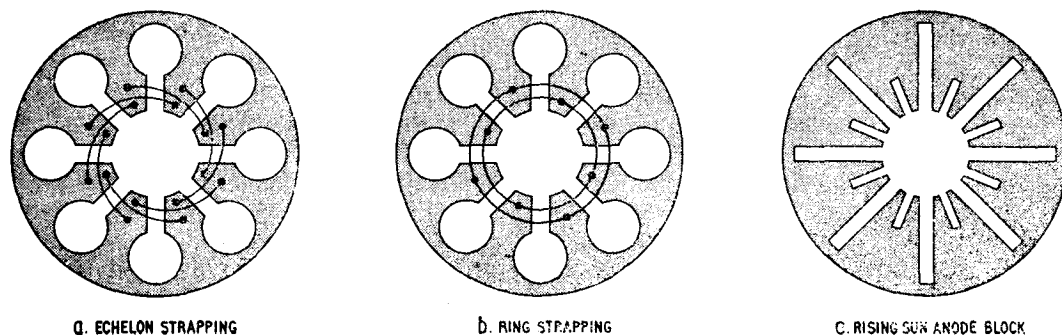


FIG. 7 METHODS OF MODE SUPPRESSION

Magnetron Performance

The output frequency of the magnetron is determined by the resonant frequency of the cavities, including the effect of the top and bottom plate spacing and by the transit time of the electrons. This, in turn, depends upon the value of the e.h.t. supply and the strength of the steady magnetic field. If either of these varies the output frequency changes. The output loop or slot is tightly coupled to the cavities and any mismatch of the load which reflects a reactive component into the cavities will "pull" the magnetron off tune. As the transit time of the electrons is no longer synchronised to the oscillatory E fields the power output falls. This effect can be caused by a fault in the output transmission system (waveguide or co-axial cable) or in the aerial system.

As the magnetron heats up, the anode block expands altering the size of the cavities and causing the frequency to drift. This effect is minimised by forced air cooling of the anode block and by automatic frequency control circuits (see Section 6).

The power output available from a magnetron depends upon the following factors.

- The power input of the modulator.
- Matching of the output to the transmission line system.
- The mode in which the magnetron oscillates.
- The strength of the H field. As the field strength of the magnet decreases with age, mishandling, etc., the anode current increases and power output falls.

The magnetron is a robust and compact source of microwave high power and is widely employed in radar transmitters. Mechanical tuning by means of flexible top and bottom end plates or by inserting a plunger into the cavities provides a narrow tuning range. Magnetrons for pulsed and c.w. use are available. Details of typical magnetrons are given in Table 1.

Type	Wavelength	Max anode Current	Voltage	Pulse duty factor	Peak Out-put Power	Efficiency
Pulsed ..	10cm	45A	25kV	·003	1MW	50%
Pulsed ..	3cm	30A	28kV	·001	300kW	40%
CW ..	10cm	130mA	1·5kV	—	50W	10%

TABLE 1. MAGNETRON OPERATING DETAILS

The M-type Backward-wave Oscillator

The tuning range of the magnetron is severely limited by the narrow bandwidth of the resonant cavities. Certain radar systems require a high-power oscillator which can be tuned over a wide frequency range without an appreciable change in power output. This need is met by the *M-type backward wave oscillator (M-carcinotron)*. As in the O-carcinotron an interdigital slow wave structure supports a backward travelling wave which is amplified by interaction with a velocity-modulated electron stream. In the M-type b.w.o. a circular slow wave structure is used and a strong magnetic field, at right angles to the steady d.c. E field, is provided by a permanent magnet.

Fig. 8 shows the essential parts of the oscillator without the magnet. The cathode emits electrons which are accelerated towards the gun anode but are deflected into the space between the sole plate and the slow wave structure by the strong d.c. H field. The electrons are attracted towards the slow wave structure which is approximately 1,000V positive relative to the sole plate, but as in the magnetron they are deflected into a series of orbits by the d.c. H field.

Thus the electrons progress towards the catcher electrode giving energy to the backward travelling wave on the interdigital structure in much the same manner as in the O-type b.w.o. Because the electron bunches travel in a series of curves they spend a relatively long time in retarding fields and thus transfer a large proportion of their energy to the backward travelling wave.

The amplified backward wave is coupled from the slow wave structure into the load and the electron circuit is completed via the catcher electrode and power supplies. As in the O-carcinotron an attenuator is necessary to suppress the forward travelling wave.

The frequency of the backward wave depends upon the mean velocity of the electron stream and this in turn depends upon the voltage difference between sole plate and slow wave structure.

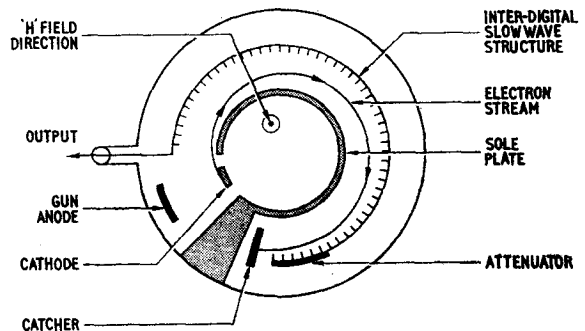


FIG. 8 M-TYPE BACKWARD WAVE OSCILLATOR

By varying this voltage difference the output frequency can be varied over a wide range.

The noise output of the M-type b.w.o. is very high over a wide frequency band and because of this it is used as a high power tunable oscillator in microwave jamming equipments (ECM equipments).

At 10cm the M-type b.w.o. can be tuned over a 2:1 frequency band and can give peak powers up to 2MW at 60 per cent efficiencies. An opened-out view of an M-type carcinotron, without the magnet, is shown in Fig. 9. It operates over the frequency range 3Gc/s to 4Gc/s with $-775V$ on the sole plate and 2500 to 4800 V on the anode.

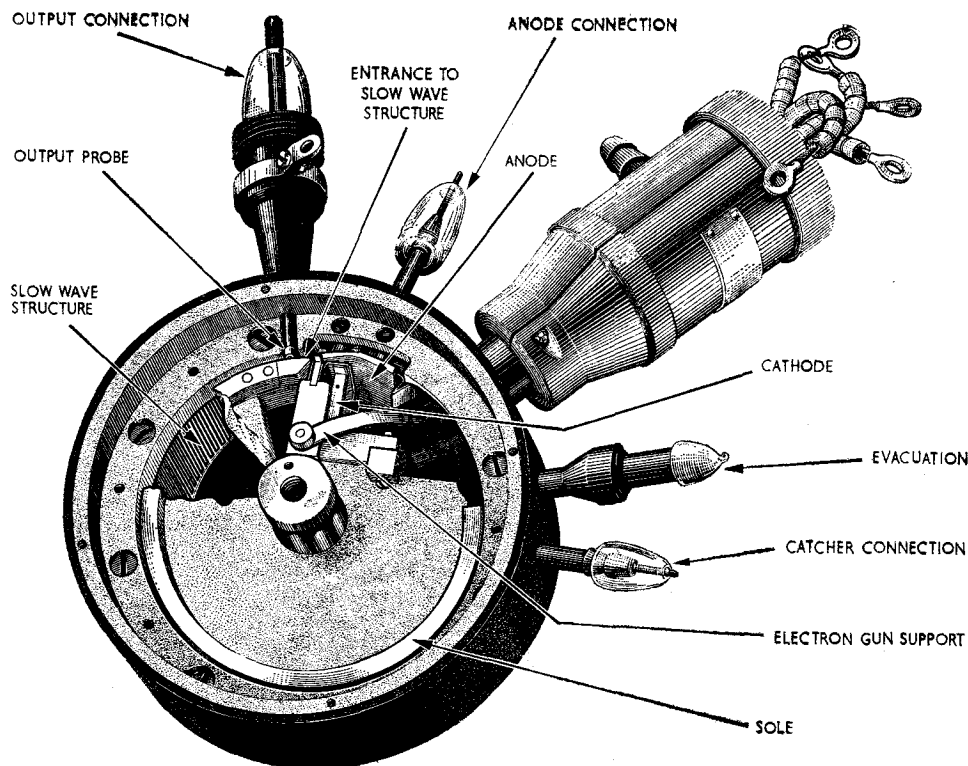


FIG. 9 M-TYPE CARCINOTRON

Platinotrons

Platinotrons are broad-band microwave M-type devices similar in appearance and action to magnetrons and M-carcinotrons. The platinotron may be used as a backward wave amplifier or as an oscillator with an external reference cavity connected to the input to control the frequency. In the former role it is called an *amplitron* and as an oscillator it is known as a *stabilotron*.

The platinotron resembles the magnetron in structure with a circular anode block into which is cut a series of vanes which form the slow wave structure. A magnetic field parallel to the central cathode is provided by a permanent magnet. The important difference between the magnetron and platinotron is that the latter has input and output couplings whereas the magnetron has only an output coupling (Fig. 10).

Amplification occurs due to the interaction between a backward travelling wave on the slow wave structure and the electron stream in much the same manner as in the M-carcinotron.

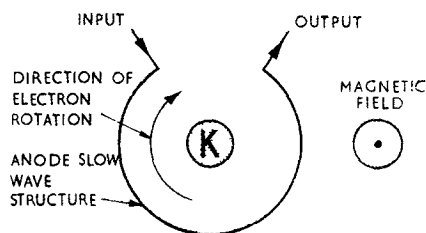


FIG. 10 THE PLATINOTRON

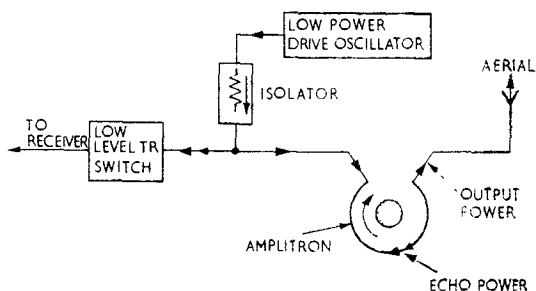


FIG. 11 USE OF THE AMPLITRON

A typical use of the amplitron in a pulsed radar system is illustrated in Fig. 11. Power from the drive oscillator is fed through an isolator to the amplitron which amplifies it to high-level power for radiation. As the power at the receiver input is small a low-level TR switch to isolate the receiver from the drive oscillator during transmission is all that is required.

The received echo travels from the aerial, round the amplitron slow wave structure, through the TR switch to the receiver input, the slow wave structure acting as a low attenuation transmission line to the echo when the modulating power is removed. Thus the advantage of using an amplitron in this system is that high level TR and TB switching is not required. The isolator between the drive oscillator and the amplitron is necessary in order to prevent reflected power due to aerial mismatch from entering the drive cavity.

The amplitron is capable of providing peak powers of up to 10MW with efficiencies of 85 per cent at L band (25cm) wavelengths. It is a broad-band device with a bandwidth of 10 per cent and a gain of about 10db. The high efficiency enables the amplitron to operate at much higher powers than klystrons or t.w.t.'s and with simpler modulating and cooling systems. No heater supplies are required. It may be used as a "booster" valve to increase the power output of an existing radar transmitter. A drawing of a complete S band (10cm) amplitron is shown in Fig. 12.

The amplitron can be employed as a very stable high-power microwave oscillator when used in conjunction with a high-Q reference cavity. The necessary feedback is provided by a mismatch unit in the output (Fig. 13). In this role it is called a stabilotron.

A portion of the amplitron output is reflected by the mismatch unit back towards the reference cavity with little attenuation. Reflected energy at the cavity resonant frequency is re-reflected in the direction in which amplification occurs in the amplitron and oscillations commence. The frequency of oscillation is determined by the reference cavity, which may be tunable. The frequency stability is up

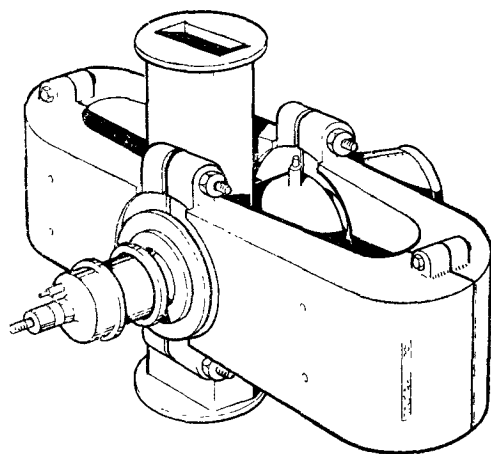


FIG. 12 AN S BAND AMPLITRON

to 100 times better than that of the magnetron and efficiencies up to 60 per cent are possible. Because of the inherent wideband properties of the platinotron the stabilotron can be tuned, by varying the frequency of the reference cavity, over a 10 per cent frequency band. Tunable stabilotrons use a variable phase shift network which adjusts the feedback phase so that oscillations are maintained.

Operating details of an amplatron and a stabilotron are given in Table 2.

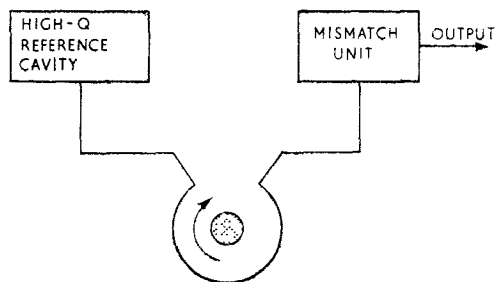


FIG. 13 THE STABILOTRON

Type	Freq Band	Freq Range (Gc/s)	Max Anode current (A)	Voltage (kV)	Peak Power Output	Gain (db)	Efficiency
Amplatron	S	2.9-3.1	180	86	10MW	8	60%
Stabilotron	L	1.2-1.4	37	38	500kW	—	45%

TABLE 2. PLATINOTRON DETAILS

CHAPTER 5

WAVEGUIDES

Introduction

In Part 1 of these notes we discussed the importance of an efficient transmission system between transmitter and aerial and between aerial and receiver. Open wire transmission lines are used to convey e.m. energy at frequencies up to about 200 Mc/s but, due mainly to radiation

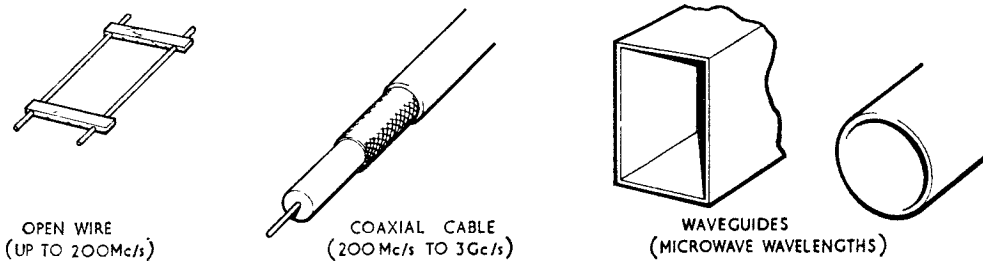


FIG 1. TRANSMISSION LINE SYSTEMS

losses, coaxial cable must be used for frequencies above 200 Mc/s up to about 3Gc/s. Above 3 Gc/s coaxial cable becomes inefficient because the skin resistance of the inner conductor results in considerable energy being lost as heat. Loss in the polythene or p.t.f.e. dielectric also becomes appreciable above 3 Gc/s (Fig. 1).

For efficient transmission of e.m. energy at centimetric wavelengths the main length of transmission line used is a hollow metal tube or *waveguide*. The e.m. wave is propagated inside the waveguide and very little radiation can occur; the large surface area of the waveguide walls result in low resistance losses; and as the energy is propagated through air inside the waveguide the dielectric losses are also low.

For these reasons waveguide is the most efficient form of transmission line at any frequency but as we shall see later the cross-sectional dimensions of the waveguide are related to the *wavelength* of the energy being propagated and at low frequencies these dimensions are usually too large for practical use.

E and H Fields

When we considered the transmission of e.m. energy along open wire and coaxial feeders we dealt in terms of voltage between the conductors and current flowing in the conductors. However, we know that where there is a voltage difference between two points there must also be an electric (E) field between those points. Similarly, a current flowing in a conductor has a magnetic (H) field around the conductor associated with it. Thus we could equally well consider the energy being propagated in terms of E and H fields and this is the method used in waveguides. A static E field can exist by itself, as can a static H field, but if an E field is *changing* it always sets up an H field and vice versa.

When we dealt with radiation from an aerial we considered e.m. energy in the form of E and H fields. We know that in free space a plane wave consists of alternating E and H fields at right angles to each other and transverse (at right angles) to the direction of propagation (Fig. 2). The wave is called a *transverse electromagnetic* (TEM) wave and moves through space with the velocity of light (c). The relative directions of the three axes can be obtained by imagining a right-hand corkscrew placed in the plane of propagation (Fig. 2a); moving the E vector towards the H vector (alphabetical order) causes the corkscrew to move in the direction of propagation.

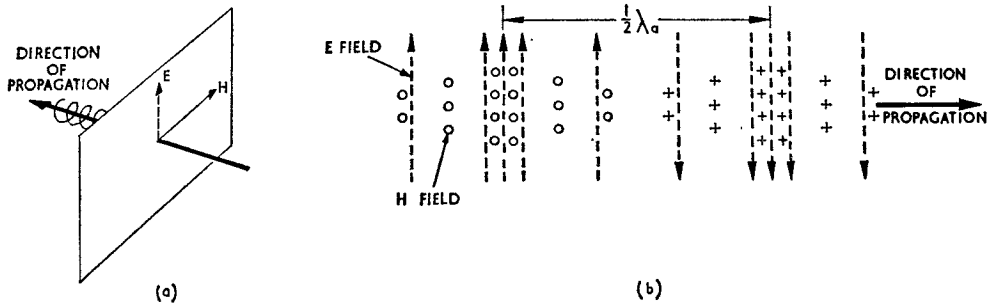


FIG 2. TRANSVERSE ELECTRO-MAGNETIC PLANE WAVE

Guided Waves

Fig. 3a shows two metal strips connected to an r.f. generator. Let us think in terms of E and H fields and see how they may be used to show how r.f. energy may be guided along the strips. If the strips are perfect conductors, i.e. if they have no resistance, the E and H fields must obey the following rules:

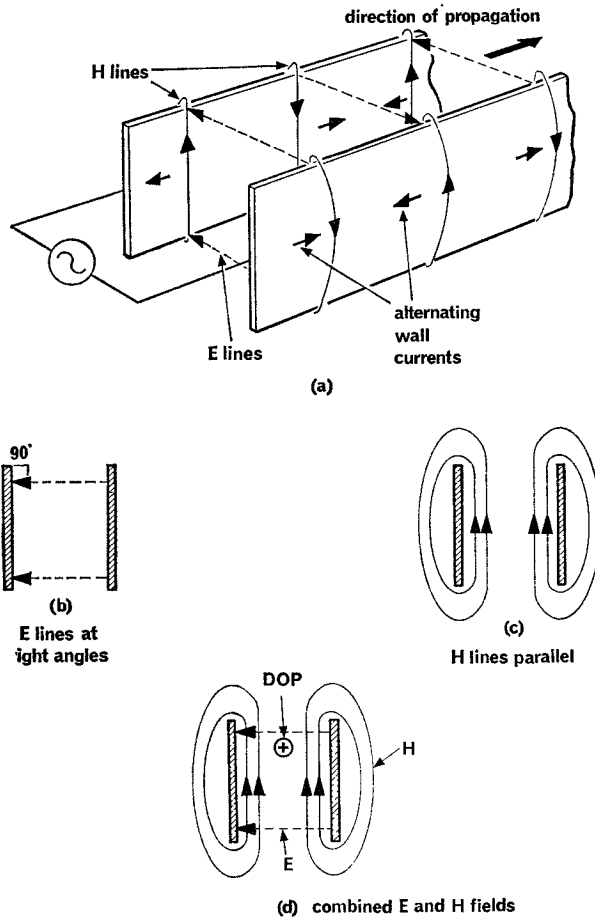


FIG 3. BOUNDARY CONDITIONS

- a. The E lines *always* terminate at 90° to the surface (Fig. 3b).
- b. The H lines *always* lie parallel to the surface of the conductor (Fig. 3c).

These rules are known as *boundary conditions*. No matter what form the fields take at a distance from the conducting surface, they must satisfy the boundary conditions *at the surface*. By applying these rules we can build up a picture of the field pattern inside any shaped waveguide. The walls of the waveguide are not perfect conductors, i.e. they have some resistance, but this affects the field patterns only slightly.

Associated with an alternating H field at a conducting surface is a current flowing on the surface. Its direction obeys the corkscrew rule and is at right angles to the H field producing it, and its magnitude depends on the strength of the H field.

Rectangular Waveguide

In order to form a rectangular waveguide and so enclose the E and H fields of the plane wave shown in Fig. 3a we shall have to add two further conducting strips AB and CD as shown in Fig. 4a. If we try to draw the E and H fields as before, we see that the boundary rules are not obeyed. The E lines terminate at 90° on conductors AC and BD but are parallel to conductors

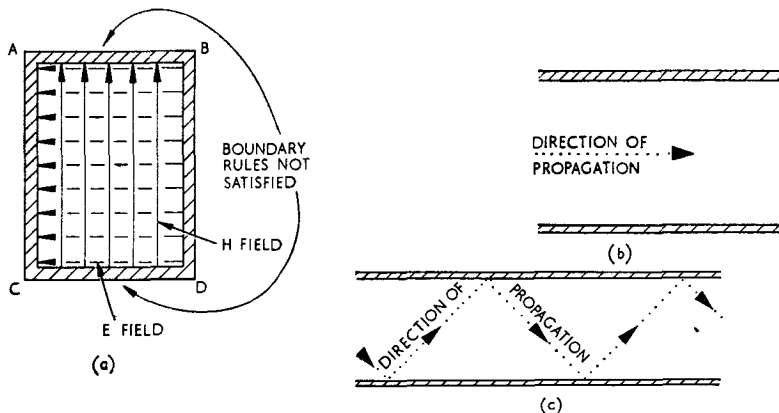


FIG. 4 PROPAGATION IN A WAVEGUIDE

AB and CD. Also the H lines, whilst parallel to AC and BD, are at right angles to AB and CD.

Thus a plane wave cannot travel directly along the axis of a waveguide (Fig. 4b). However, if the wave front moves at an *angle* to the top and bottom walls of the waveguide (Fig. 4c) a field pattern which *does* obey the boundary rules is produced and energy is propagated down the waveguide (Fig. 5). We shall now consider how this field pattern is formed.

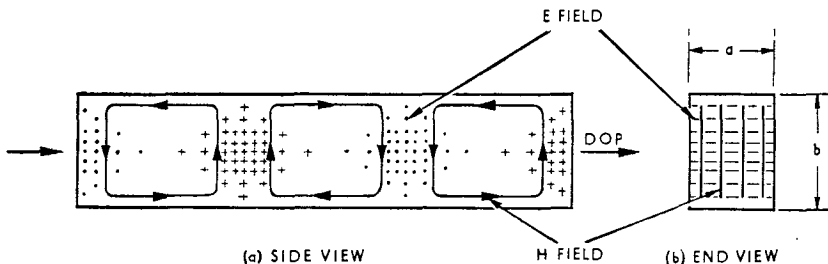


FIG. 5 FIELD PATTERN IN A RECTANGULAR WAVEGUIDE

Development of Rectangular Guide Field Pattern

In Fig. 6a an e.m. wave is shown striking a horizontal conducting surface at an angle θ . The E field is parallel to the conducting surface and therefore the E field maximum is shown going into and out of the paper with phase reversals every half wavelength ($\frac{1}{2}\lambda_a$). The H field maximum

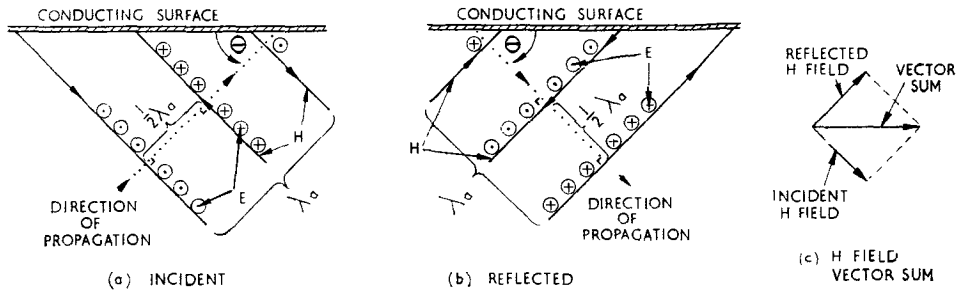


FIG. 6 INCIDENT AND REFLECTED WAVES

is represented by full lines at right angles to the direction of propagation, the arrows showing the reversals of phase every half wavelength. For clarity, only the maximum E and H fields are shown although, of course, the fields will vary sinusoidally. The direction of propagation is shown by the dotted line.

When the wave front strikes the conducting surface it is reflected, the new direction of propagation making the same angle θ with the reflecting surface as did the incident wave. The boundary condition is that at the surface the E field cannot exist parallel to the surface—it must be perpendicular or not exist at all. Therefore, since there is no perpendicular component, the E field must be zero at the surface. To satisfy this requirement an E field of equal amplitude but opposite phase to the incident E field is reflected from the surface.

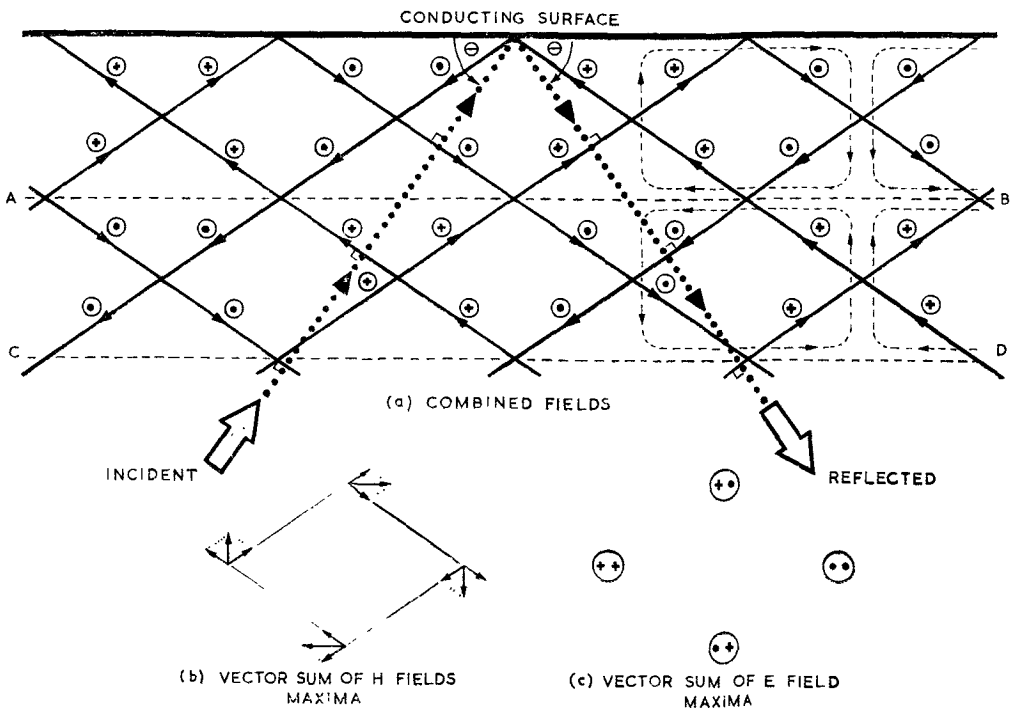


FIG. 7 COMBINED INCIDENT AND REFLECTED FIELDS

The boundary condition for the H field is that at the surface it must lie parallel to the surface, with no perpendicular component. Hence on reflection the H field suffers a 20° change in direction and its phase reverses. The vector sum of the incident and reflected H fields lies parallel to the conducting surface at the surface (Fig. 6c). Thus boundary conditions for the combined incident and reflected waves are satisfied for both E and H fields. Adding vectorially the fields at all points (Fig. 7) gives H field loops and a diamond pattern of E fields as shown in Fig. 8.

In addition to satisfying boundary conditions at the conducting surface, the rules are also satisfied at planes AB and CD . Therefore a second conducting surface may be inserted at one of these planes, for example at plane AB . This gives rise to further reflections and the incident wave will now be propagated by "bouncing" between the two surfaces; the field pattern will be that shown in the top portion of Fig. 8. The resultant direction of propagation is parallel to the reflecting walls.

Had the lower reflecting surface been placed at plane CD instead of at AB the pattern would contain *two* H loops, one above the other and energy is said to be propagated in a higher *mode*.

To confine the wave pattern in the horizontal plane a pair of plates is placed at right angles to the reflecting surfaces and a rectangular waveguide is thus formed (Fig. 9b). These plates satisfy boundary conditions; their minimum distance apart depends upon the power to be

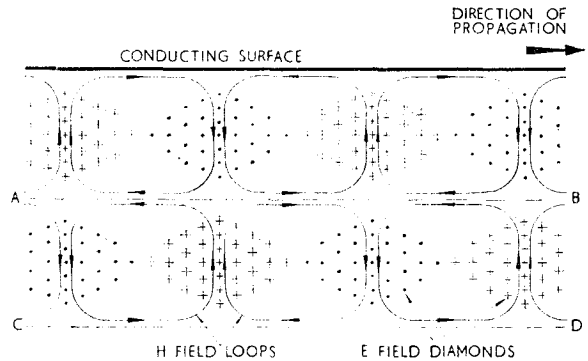


FIG. 8 RESULTANT FIELD PATTERN

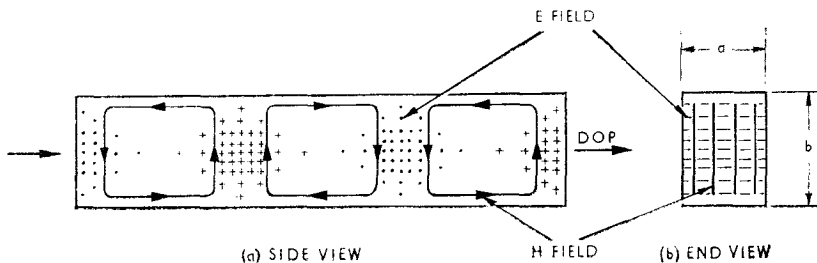


FIG. 9 THE H_{01} MODE IN A RECTANGULAR GUIDE

propagated since if they are too close the resultant concentrated E field may arc between the plates.

The narrow dimension of the waveguide is usually called the a dimension and the broad side the b dimension.

So far we have seen that:

- A plane wave cannot go directly down a waveguide since it does not satisfy boundary conditions.
- A wave which "bounces" can move down a waveguide because the overlapping field patterns of the incident and reflected waves combine to produce a pattern that *does* obey boundary conditions.
- This new pattern is not a TEM wave.

The field pattern formed inside the waveguide is shown in Fig. 9. Since there is a component of the H field in the direction of propagation it is called an H mode. It is sometimes called a *transverse electric* (TE) mode since the E field is transverse to the direction of propagation. To

to distinguish this mode from other H modes it is known as an H_{01} mode. The subscript 0 indicates that there is no change in field strength across the narrow side of the guide and the subscript 1 indicates that the field goes through one maximum value as we move across the broad side of the guide.

The field pattern shown in Fig. 9 is that normally used to propagate e.m. energy down a rectangular waveguide and the following differences between this H_{01} guided wave and free space propagation should be noted. The length of waveguide occupied by a complete cycle of wave pattern (two H loops) is called a *guide wavelength*, λ_g (Fig. 10). This is always *greater* than a free space wavelength (λ_a). The relation between λ_g and λ_a depends upon the *broad* (b) dimension of the waveguide but is independent of the narrow (a) dimension. As we shall see later, the free space wavelength must always be less than

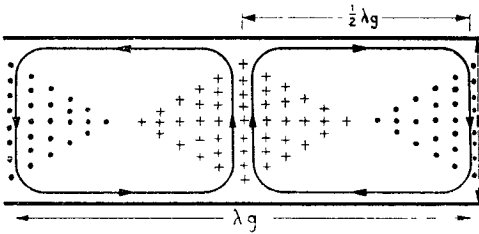


FIG. 10 GUIDE WAVELENGTH

twice the broad dimension otherwise the wave is not propagated. The *critical* or *cut-off* wavelength for which $\lambda_a = 2b$ is denoted by λ_c .

The velocity with which the wave *pattern* moves down the guide is called the *phase velocity* (v_p) and is always *greater* than the velocity of light (c). The component of the wave velocity along the axis of the guide is the *group velocity* (v_g) and is the velocity with which the energy actually travels down the guide; it is always *less* than the velocity of light.

λ_a , λ_g and Broad Dimension

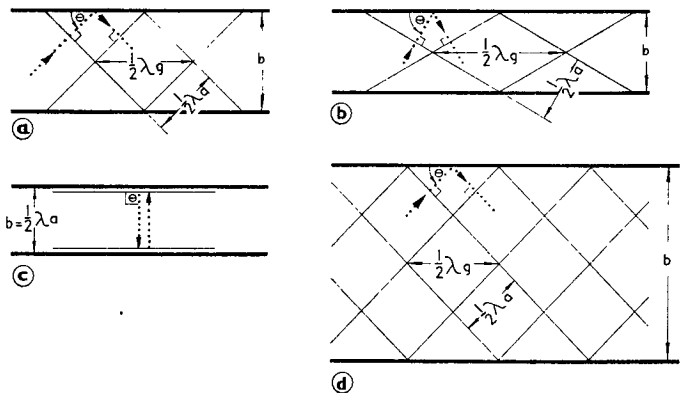
Fig. 11a represents a section of the field pattern of Fig. 7a. For clarity, only the incident and reflected H field maximum are shown; the E field and resultant H loops are not shown.

In Fig. 11b the broad (b) dimension has been reduced and λ_a remains unchanged. In order that boundary conditions are satisfied the angle θ increases; the diamond-shaped H lines become "squashed out" and $\frac{1}{2}\lambda_g$ *increases*.

Fig. 11c shows the b dimension further reduced until it equals $\frac{1}{2}\lambda_a$. The angle θ becomes 90° and the wave front becomes parallel to top and bottom guide walls. The wave front bounces between these two walls and no axial propagation occurs, i.e. cut-off has been reached.

In Fig. 11d the b dimension is greater than λ_a . This allows two H loops to form and the wave is propagated in the higher H_{02} mode. Cut-off wavelength for this mode occurs when $\lambda_a = b$.

Thus to propagate energy in the H_{01} mode the waveguide must be designed with a b dimension *between* $\frac{1}{2}\lambda_a$ and λ_a . The b dimension is usually made $0.707\lambda_a$ as at this size the waveguide losses are lowest for propagating the H_{01} mode. The narrow a dimension must be less than $\frac{1}{2}\lambda_a$ to prevent the field pattern moving through 90° into the other plane, i.e. to prevent the

FIG. 11 EFFECT OF ALTERING BROAD DIMENSION WITH λ_a KEPT CONSTANT

pattern shifting to the other walls so that a becomes b and vice versa. However, the a dimension must not be made too narrow otherwise arcing between the walls may result.

The normal waveguide is designed to propagate energy of a certain free space wavelength and its dimensions are fixed. However, similar effects to those illustrated in Fig. 11 are obtained by variations in λ_a . Thus for the fixed b dimension of Fig. 11a an increase in λ_a (decrease in frequency) would result in an increased λ_g as in Fig. 11b. If λ_a is increased until it equals $2b$, cut-off is reached as in Fig. 11c. Similarly, if λ_a is decreased (frequency increased) such that λ_a is less than the b dimension the H_{02} mode can propagate as in Fig. 11d. Small variations in λ_a do not affect wave propagation greatly.

The dimensions of typical standard British rectangular waveguides for propagation in the H_{01} mode are given in Table 1. Also shown is the db attenuation per 100 feet length of guide.

Phase and Group Velocities

In a rectangular waveguide the H lines (AB in Fig. 12) move at an angle of incidence ϕ to the guide walls with the velocity of light c . Thus in the time that point B moves to D, point A moves *along the guide wall* to D. As the distance AD is greater than the distance BD the point A moves along the guide wall at a velocity *greater* than that of light. This is the *phase velocity*, v_p , of the wave in the guide.

Band	Freq.	λ_a	Internal dimensions		Attenuation db/100 ft.
			Inches	Centimetres	
S	3 Gc/s	10 cm	2.84 x 1.34	7.2 x 3.4	0.55
X	10 Gc/s	3 cm	0.90 x 0.40	2.3 x 1	3.24

TABLE 1. WAVEGUIDE DIMENSIONS AND ATTENUATION

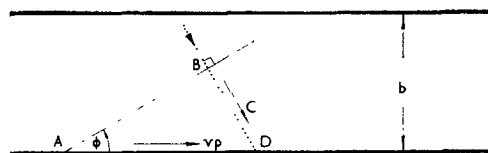


FIG. 12 PHASE VELOCITY

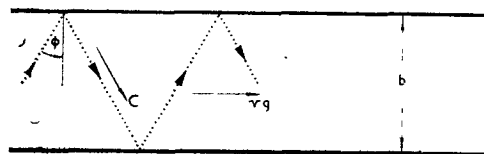


FIG. 13 GROUP VELOCITY

It should be noted that the phase velocity does not represent any physical movement down the guide, but is merely the movement of a point at which a phase change occurs at the guide wall.

Since the e.m. energy is following a zig-zag path down the waveguide with a velocity c (Fig. 13) the resultant velocity *along the axis* of the guide must be *less* than c . This resultant velocity is the *group velocity*, v_g , of the wave in the guide.

We know that for a wave of frequency f c/s in free space, $c = f\lambda_a$. The corresponding relationship for a guided wave is $v_p = f\lambda_g$. Since v_p is always greater than c , λ_g must be greater than λ_a (Fig. 14).

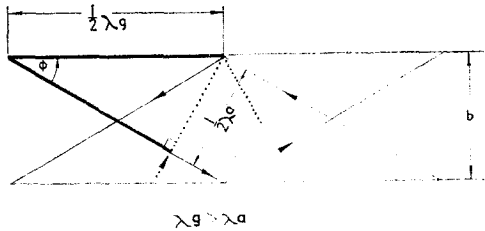


FIG. 14 WAVEGUIDE RELATIONSHIPS

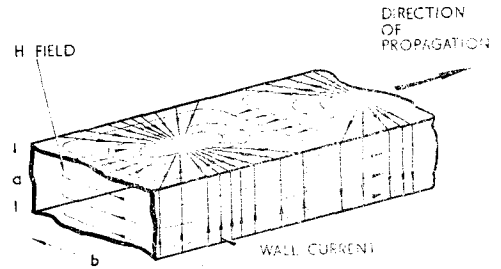


FIG. 15 WALL CURRENTS

Waveguide Losses

An alternating magnetic field parallel to a conductor induces a current in the conductor. The direction of current flow is at right angles to that of the field causing it, and the amplitude of the current depends upon the intensity of the field. In waveguides these currents are confined to a very thin skin on the inside surface of the walls. Thus in a rectangular waveguide propagating an H_{01} wave the wall current pattern is as shown in Fig. 15. This is an instantaneous picture of the wall currents; they are, of course, alternating and change direction every half cycle so they do not build up on the waveguide walls. It is rather like alternating current flowing into a capacitor. The whole wall current pattern moves along the guide walls at the phase velocity of the wave.

Since the waveguide walls possess some resistance, power is lost as the wave travels down the guide. Attenuation in a rectangular guide depends upon:—

- Guide dimensions.
- Resistance of the walls.
- Frequency.
- Mode of propagation.

Since attenuation depends upon surface resistance it is important to avoid corrosion on the inner surfaces. For this reason waveguides are often silver or cadmium-plated on the inside walls or are sealed off at each end by plastic diaphragms.

Waveguide Characteristic Impedance

In transmission line theory the characteristic impedance of a line is defined as the ratio of voltage to current at any point on an infinite line. For a plane wave in free space the corresponding ratio is that of the E field strength to the H field strength at any point where the E and H fields are at right angles and are transverse to the direction of propagation. This ratio is called the *characteristic impedance of free space* (Z_w) and has a value of 377 ohms.

The characteristic impedance of a waveguide (Z_H) propagating a wave in the H_{01} mode can be calculated in a similar manner. The relationship is:

$$Z_H = 377 \frac{\lambda_g}{\lambda_a} \text{ ohms.}$$

Since λ_g is always greater than λ_a , Z_H is greater than Z_w . If the b dimension is gradually increased, λ_g becomes smaller and the guide impedance falls towards that of free space.

Circular Waveguide

Rectangular waveguides propagating the H_{01} mode are used for the main lengths of waveguide runs in most radar systems. However, short lengths of circular waveguide are required

in systems which use a rotating aerial. The E and H fields inside a circular guide must also obey boundary rules and if these rules are applied we can build up the field pattern inside a circular waveguide.

Fig. 16a shows the field pattern inside a length of coaxial cable. Note that boundary rules are satisfied for both inner and outer conductors. Fig. 16b shows how this pattern alters to obey boundary rules when the centre conductor is removed and a circular waveguide formed. This field pattern is called the E_{01} mode; E indicates that a component of the E field is parallel to the direction of propagation; the first subscript, nought, indicates the number of *complete* wavelength changes in field pattern moving around the *circumference*; and the second subscript, one, indicates the number of *half-wavelength* changes moving across a *diameter*. The mode is sometimes called a *transverse magnetic* (TM) mode.

The E_{01} mode is not the only field pattern which can form in a circular guide but it is the one most suitable for use with a rotating aerial since it has circular symmetry. The H_{11} mode shown in Fig. 16c does not have circular symmetry and although this is the dominant mode for circular waveguide, i.e. it has the longest cut-off wavelength, it is not suitable for use with a rotating device. Therefore steps are usually taken to stop the H_{11} mode forming, and to encourage the formation of the E_{01} mode.

The cut-off wavelength for a circular guide is proportional to the diameter of the guide; guide wavelength (λ_g) is always greater than free space wavelength (λ_a), just as in a rectangular guide; the characteristic impedance of a circular guide propagating the E_{01} mode is less than that of free space and is given by $Z_E = 377 \frac{\lambda_a}{\lambda_g}$ ohms.

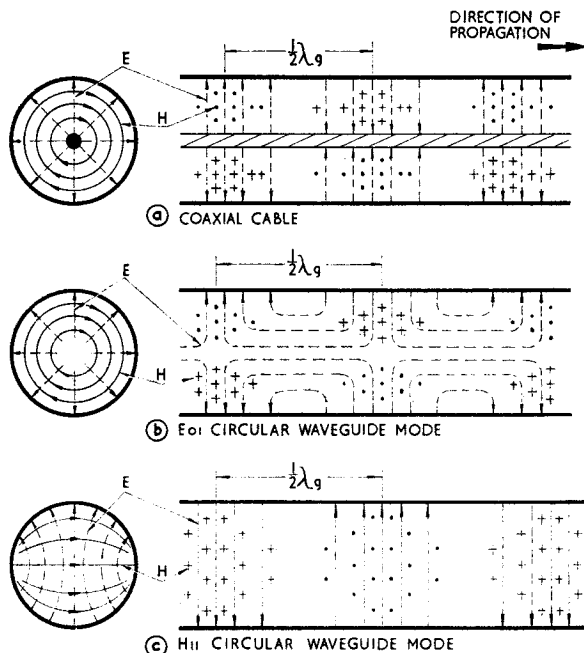


FIG. 16 COAXIAL CABLE AND CIRCULAR WAVEGUIDE FIELD PATTERNS

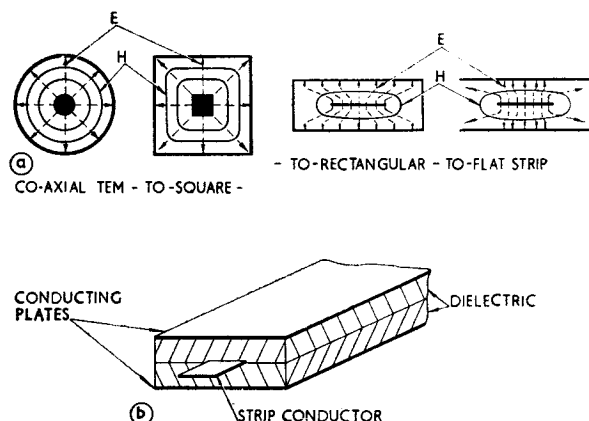


FIG. 17 FORMATION OF TEM MODE IN A FLAT-STRIP TRANSMISSION LINE

Microwave Printed Circuits

An important development in transmission of microwave energy is that of *microwave printed circuit* (m.p.c.) or *flat-strip* components. Normal printed circuit techniques are used to manufacture these smaller and lighter alternatives to waveguides.

Fig. 17 shows how the field pattern inside a coaxial transmission line adapts to form the TEM mode in a three-conductor flat-strip line. Almost all the E field is concentrated

(A.L. 2, August, 1965)

near the centre strip. Since there is no voltage between the two outer conductors the fringe field decreases rapidly away from the centre conductor. Thus no side walls are required.

The three-conductor flat-strip transmission line (sometimes called septate or tri-plate) is the most commonly used m.p.c. line and consists of a printed or metal foil centre conductor and two dielectric spacers between two outer conductors (Fig. 17b).

Another form of m.p.c. transmission line evolved from its twin wire counterpart is shown in Fig. 18. The field is contained between the strip conductor and the ground plane. It is an unshielded system and it is therefore more difficult to confine all the energy in the vicinity of the strip.

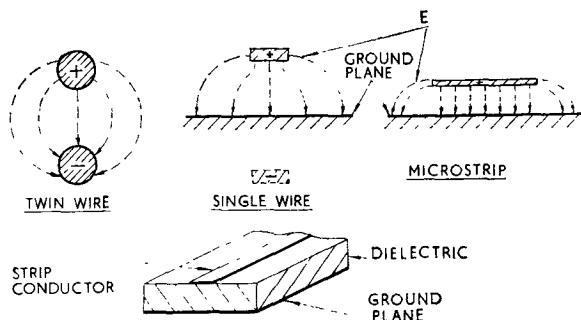


FIG. 18 MICROSTRIP LINE

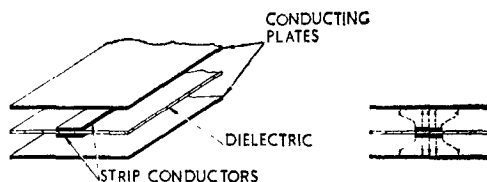


FIG. 19 FOUR-CONDUCTOR STRIP-LINE

A third form of m.p.c. transmission line is shown in Fig. 19. This is used where dielectric losses and weight must be small or where high power is to be carried.

The characteristic impedance of m.p.c. transmission lines depends upon the width of the narrow strip and so they are easily matched into coaxial connectors.

The required mode in flat-strip transmission line is the TEM coaxial mode and to prevent higher modes forming, the ground plane spacing should be less than $\frac{1}{2}\lambda$. Higher modes may form due to a non-symmetrical junction with coaxial line or longitudinal tilting of the centre strip. These unwanted modes can be eliminated by placing shorting pins between the outer conductors.

CHAPTER 6

WAVEGUIDE COMPONENTS

Introduction

A waveguide designed to convey e.m. energy in a radar system is not a simple straight length of rectangular guide. It contains many components which increase its usefulness and efficiency. These components together with the waveguide form a waveguide system, and in this chapter we shall consider the function of some common waveguide components.

Coupling Devices

Electromagnetic energy is introduced into a waveguide in such a way that the E and H fields from the coupling device form the desired mode inside the guide, e.g. to form the H_{01} mode in a rectangular guide the coupler is positioned so that the E field is launched perpendicular to the broad sides. A suitably positioned launching device will also extract energy from the guide.

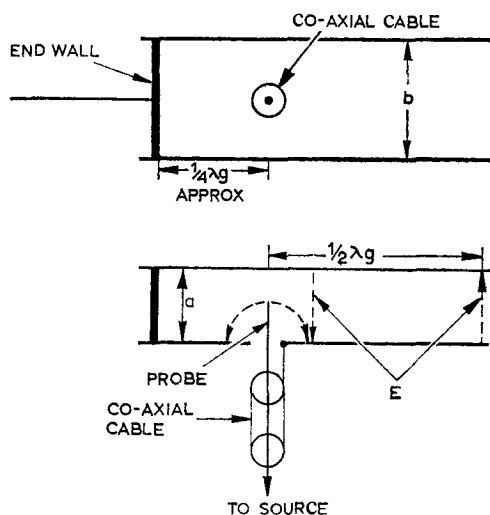


FIG. 1 PROBE COUPLING

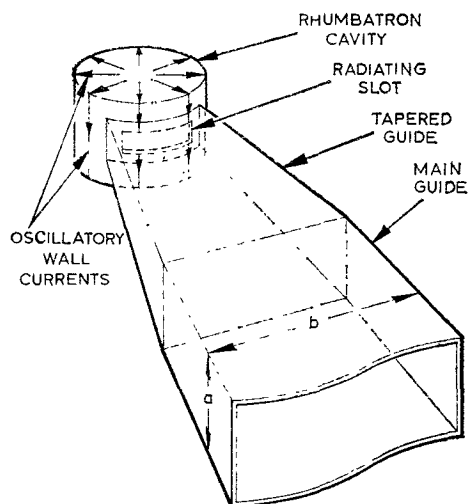


FIG. 2 SLOT COUPLING FROM CAVITY TO RECTANGULAR WAVEGUIDE

A common coupling device is shown in Fig. 1. Energy is conveyed from source to waveguide by a short length of coaxial cable the extended centre conductor of which forms a *probe* projecting through the centre of the broad side of the waveguide. The probe acts as a small aerial inside the guide and for maximum coupling the probe must be correctly matched to the coaxial cable and waveguide. This is achieved by adjusting the length of the probe and its distance from the end wall. By mounting the probe in the centre of the broad side the E field is radiated perpendicular to the broad sides and the desired H_{01} mode is formed.

The probe may be used to extract energy from the waveguide since E and H fields in the guide will induce voltages and currents into the probe causing current to flow in the coaxial cable.

Loop coupling is sometimes employed in waveguides. The loop projects from the narrow dimension so that the H lines thread the loop. The amount of coupling may be varied by rotating the loop.

Energy may also be launched into or extracted from a waveguide by *coupling slots*. We saw in Chapter 1 (p.252) that a slot cut in the wall of a resonant cavity such that it interrupts the flow of wall current will radiate e.m. energy. Fig. 2 shows how energy radiated by such a slot may be

conveyed from the cavity by a waveguide. The slot impedance is matched to the rectangular guide by a tapered section guide. The planes of the fields radiated from the slot are the same as those of the required guide mode. For maximum coupling the slot must interrupt the wall current at right angles as in Fig. 2. If the slot is tilted relative to the wall current the coupling is reduced.

Fig. 3 shows wall currents in an H_{01} rectangular guide with slots cut in the broad and narrow sides. Slots 1 and 2 interrupt the wall currents at right angles and so maximum coupling occurs from these slots.

Slots 3 and 4 are parallel to the wall currents and no energy is coupled from these slots. They may, however, be used to allow a probe to be inserted into the guide to take measurements.

Slot 5 is inclined to the wall current and slot 6 is displaced from the centre of the broad dimension and interrupts the current; both these slots will couple energy from the guide but the degree of coupling is less than for slots 1 and 2 and varies with the inclination or displacement. For least disruption of the waveguide field pattern the slots should be a half wavelength long.

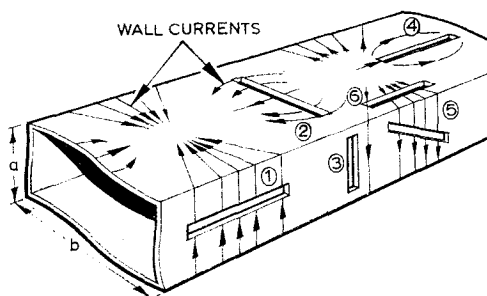


FIG. 3 SLOTS IN A RECTANGULAR WAVEGUIDE

Waveguide Stubs

When a waveguide is terminated with a short-circuit, reflection occurs and the combination of reflected and incident waves forms standing waves inside the guide. Resonant sections of waveguide so formed have properties and uses similar to their twin transmission line counterparts.

Fig. 4 shows a short-circuited shunt stub which projects from the narrow side of the main guide. The broken lines indicate the twin transmission line equivalent. If the stub length (X),

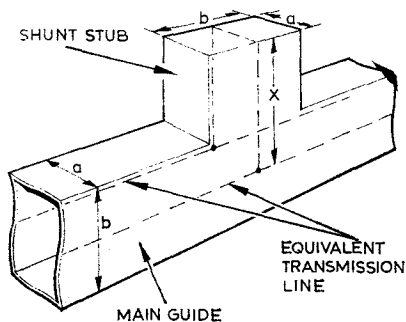


FIG. 4 SHUNT STUB

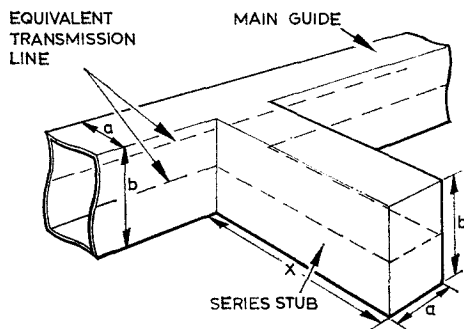


FIG. 5 SERIES STUB

measured from the centre of the broad side of the main guide, is an *even* number of quarter guide wavelengths ($\frac{1}{2}\lambda_g$) the short-circuit termination is reflected as a short-circuit across the main guide and no energy passes the stub.

If the stub length is an *odd* number of $\frac{1}{4}\lambda_g$ the short-circuit termination is reflected as an open-circuit across the main guide and the stub acts as a metallic insulator, i.e. it has no effect on the transmission of energy down the guide.

Intermediate stub lengths reflect either inductive or capacitive reactance across the main guide and can be used as matching devices. Stubs with open-circuit terminations are not used because radiation occurs.

A short-circuit series stub projecting from the *broad* side of the main waveguide is shown in Fig. 5. With this stub the short-circuited termination is reflected over an *even* number of $\frac{1}{4}\lambda_g$ as a short-circuit and completes the guide wall allowing energy to flow down the main guide. If the stub is an *odd* number of $\frac{1}{4}\lambda_g$ long, measured from the guide wall, energy cannot propagate.

Thus note that a $\frac{1}{4}\lambda_g$ short-circuited *shunt* stub placed in the *narrow* wall of the guide appears as an open-circuit *across* the guide and allows energy to pass normally, whereas a similar stub placed in the *broad* wall of the guide (a *series* stub) effectively open-circuits the broad wall and prevents the propagation of energy. For $\frac{1}{2}\lambda_g$ short-circuited stubs the results are reversed—in the shunt case no energy passes down the guide; in the series case energy is propagated normally.

Either shunt or series stubs may be used to interrupt or complete the main length of guide. Stub lengths can be adjusted by a tuning plunger to reflect a suitable reactance for matching purposes.

Waveguide Irises

Most waveguide components introduce a certain amount of mismatch into the main waveguide. This mismatch results in standing waves forming in the guide; the voltage maxima can cause arcing and the current maxima can increase waveguide losses. To avoid these undesirable effects we introduce a reactance to cancel the unwanted mismatch. As already indicated, this can be done by stubs as in conventional transmission lines; an alternative in waveguides is the *matching iris*.

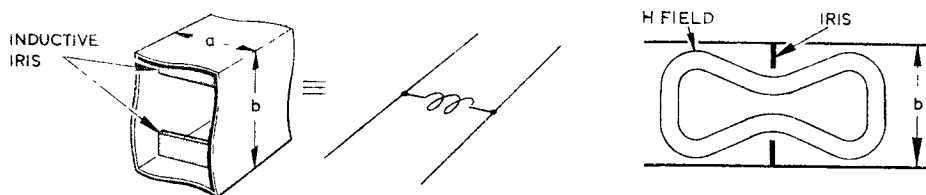


FIG. 6 INDUCTIVE IRIS

If the broad dimension of the guide is restricted by metal plates as shown in Fig. 6 the H field is modified and an *inductive* reactance is introduced into the guide.

In Fig. 7 the narrow dimension is restricted by metal plates and the intensity of the E field is increased at this point. Thus a *capacitive* reactance is introduced into the guide. Use is restricted by the risk of arcing at high power. An alternative is to insert a low-loss dielectric material into the guide.

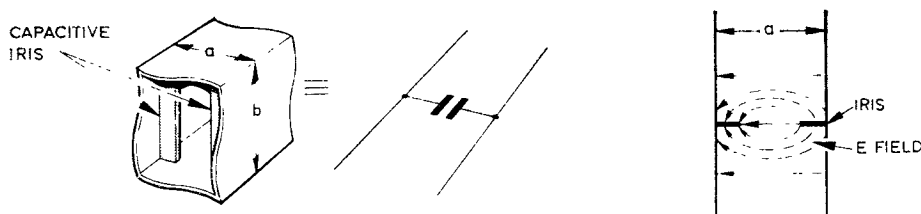


FIG. 7 CAPACITIVE IRIS

A combination of inductive and capacitive irises forms a *resonant iris* at a desired frequency (Fig. 8). At the resonant frequency it has no effect, i.e. it is transparent to energy in the guide; to frequencies above resonance it is capacitive; to frequencies below resonance it is inductive. The waveguide can be sealed against moisture or pressure variations by a glass or mica window mounted in the resonant iris.

RESONANT IRIS

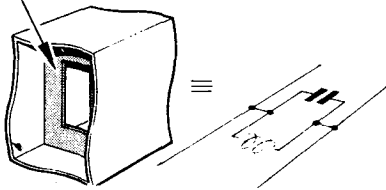


FIG. 8 RESONANT IRIS

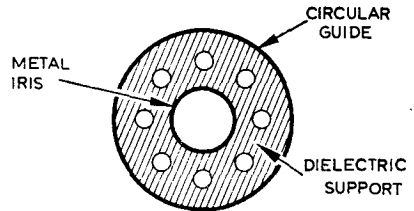


FIG. 9 RING IRIS

Fig. 9 shows a form of iris often used in circular waveguides to suppress the H_{11} mode so that the symmetrical E_{01} mode may form instead. A metal ring is mounted concentric with the guide. Since boundary rules are not satisfied for the H_{11} mode the iris is opaque to this mode but transparent to the E_{01} mode where boundary rules are satisfied.

A metal screw inserted through the centre of the broad dimension acts as a variable matching device (Fig. 10). It is resonant at a certain length and acts as a series tuned circuit, shorting the waveguide and preventing energy passing, i.e. it is opaque to energy at a certain frequency. With reduced length it is capacitive.

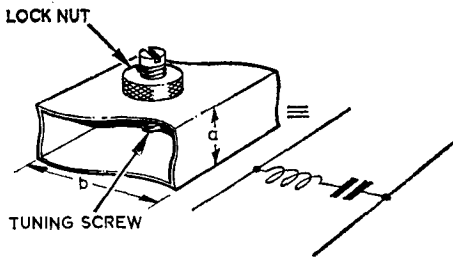


FIG. 10 TUNING SCREW

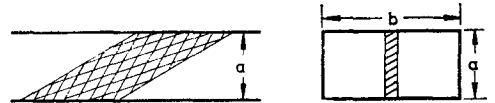


FIG. 11 TAPERED VANE FIXED ATTENUATOR

Attenuators

It is often necessary, e.g. when making measurements, to attenuate the circuit power. The waveguide equivalents of the resistor and potentiometer are the fixed and variable attenuator. In general, a waveguide attenuator consists of a shaped mounting coated with a thin film of resistive substance such as iron dust or carbon, placed so that part of the waveguide energy is absorbed by the resistive element.

The fixed attenuator shown in Fig. 11 consists of a tapered vane coated with a resistive material and mounted in the guide as shown. If the taper is gradual and takes place over several guide wavelengths, reflection is negligible and part of the guide energy is absorbed. Attenuation is maximum when the vane is placed in the centre of the broad dimension as shown and decreases when placed close to the wall. When suitably mounted the tapered vane can be made into a variable attenuator.

A spiral-shaped resistive vane projecting through a slot along the centre of the broad side acts as a variable attenuator (Fig. 12). Attenuation is proportional to the depth the vane penetrates into the guide and this is made variable by a rotatable mounting which may have a micrometer adjustment. Radiation losses are prevented by enclosing the moving parts in a metal casing. As the vane enters the field gradually, minimum reflection occurs and the standing

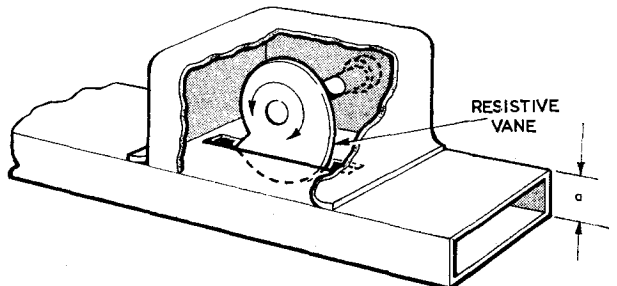


FIG. 12 VARIABLE ATTENUATOR

wave ratio due to the attenuator may be as low as 1:01:1.

A wedge of resistive-coated material mounted at the end of a waveguide as shown in Fig. 13 completely absorbs the incident energy without reflections and thus acts as a matched termina-

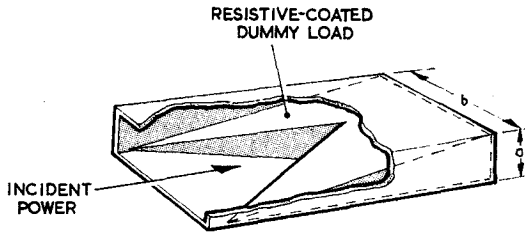


FIG. 13 TAPERED DUMMY LOAD

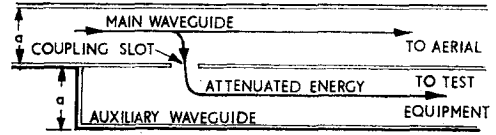


FIG. 14 SLOT-COUPLED ATTENUATOR

tion or dummy load.

In radar equipments it is often necessary to sample the transmitted energy to check frequency, p.r.f. and power output. In Fig. 14 two sections of waveguide have a common wall in which a coupling slot or hole is cut. The proportion of energy fed from the main guide to the auxiliary waveguide is determined by the shape and position of the coupling slot. This is set at manufacture to provide the required amount of attenuation.

Waveguide Joints

Sections of waveguide are manufactured in convenient lengths which can be joined together. A common method of jointing is by terminating each section of guide with flat plates called *flanges*. The flanges are then bolted together (Fig. 15). To prevent reflections from this type of joint the flanges must be perfectly flat and at right angles to the guide. A small gap between the sections could cause standing waves in the guide and arcing at high powers. Thus a high degree of manufacturing precision is required with consequent high cost.

This accuracy can be dispensed with by using the *choke flange* joint shown in Fig. 16. It consists of a normal flange on one waveguide section screwed on to an enlarged flange in which a groove $\frac{1}{4}\lambda_g$ deep is cut as shown. The short circuit at A is reflected as an open-circuit at B. This is reflected over a further $\frac{1}{4}\lambda_g$ as a short

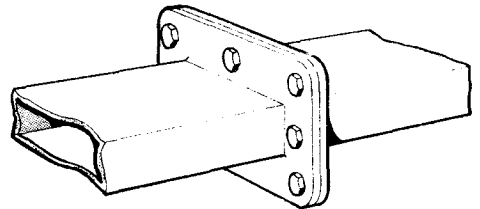


FIG. 15 FLANGE JOINT

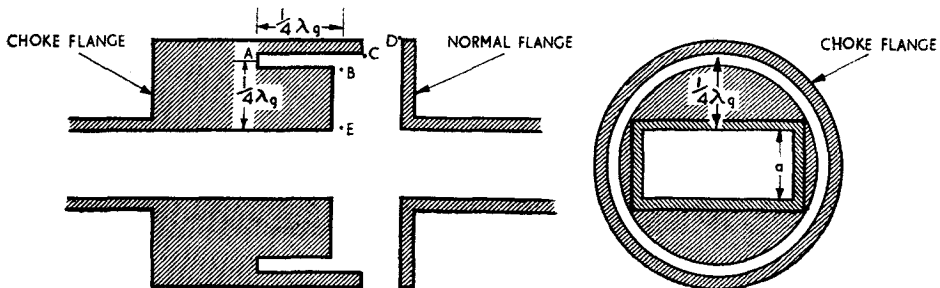


FIG. 16 CHOKE FLANGE JOINT

circuit at E so electrically completing the guide walls. Thus although the mechanical joint between the two flanges may not be perfect, no radiation, reflection or arcing can occur.

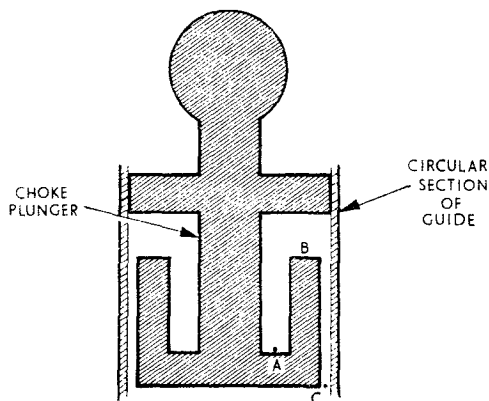


FIG. 17 PLUNGER CHOKE JOINT

A further application of the choke joint is illustrated in Fig. 17. The lower end of a circular waveguide tuning plunger makes an electrical short-circuit with the guide walls although there is little mechanical contact. Since $AB = BC = \frac{1}{4}\lambda_g$ the short-circuit at A is reflected over twice $\frac{1}{4}\lambda_g$ as an electrical short-circuit at C.

A common requirement in radar equipments is for a joint between a fixed section of guide and a section feeding a rotating aerial. The main length of rectangular guide propagating the H_{01} mode is joined to a short section of circular guide in which the E_{01} mode is propagated. The circular guide is then converted back to rectangular for the remainder of the run to the aerial. A rotating joint is included in the circular section. The electrical details are shown in Fig. 18.

The H_{01} mode in the input rectangular guide is transformed into the E_{01} circular mode as shown in Fig. 18a. Similarly, the axial E field in the circular section launches an H_{01} mode into

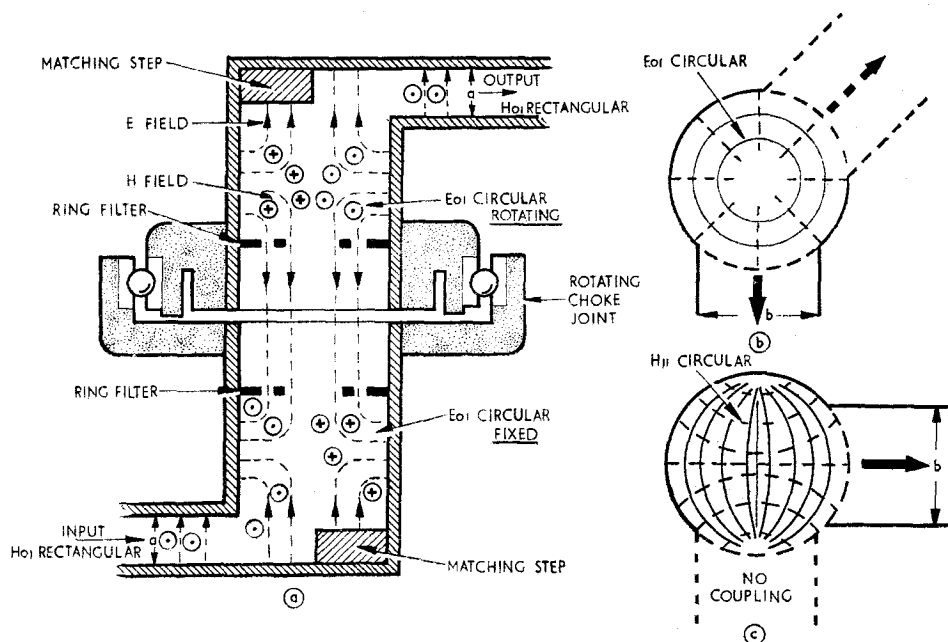


FIG. 18 THE ROTATING JOINT

the rectangular output guide. Because of the axial symmetry the E_{01} circular mode will present the same field pattern to the output rectangular guide in whatever direction it may be turned. This is not so if the H_{11} circular mode is allowed to form (Fig. 18 *b* and *c*).

To block the unwanted H_{11} circular mode two ring filters (see p.292) are mounted as shown in Fig. 18*a*. To match the rectangular and circular guide impedances matching steps or posts are employed. The union between the fixed and rotating sections employs a choke joint.

Bends, Corners and Twists

To change the direction of a guide run, bends and corners may be used. Bends and corners in the planes of the H and E fields are shown in Fig. 19*a* and *b*. The bend or corner causes some

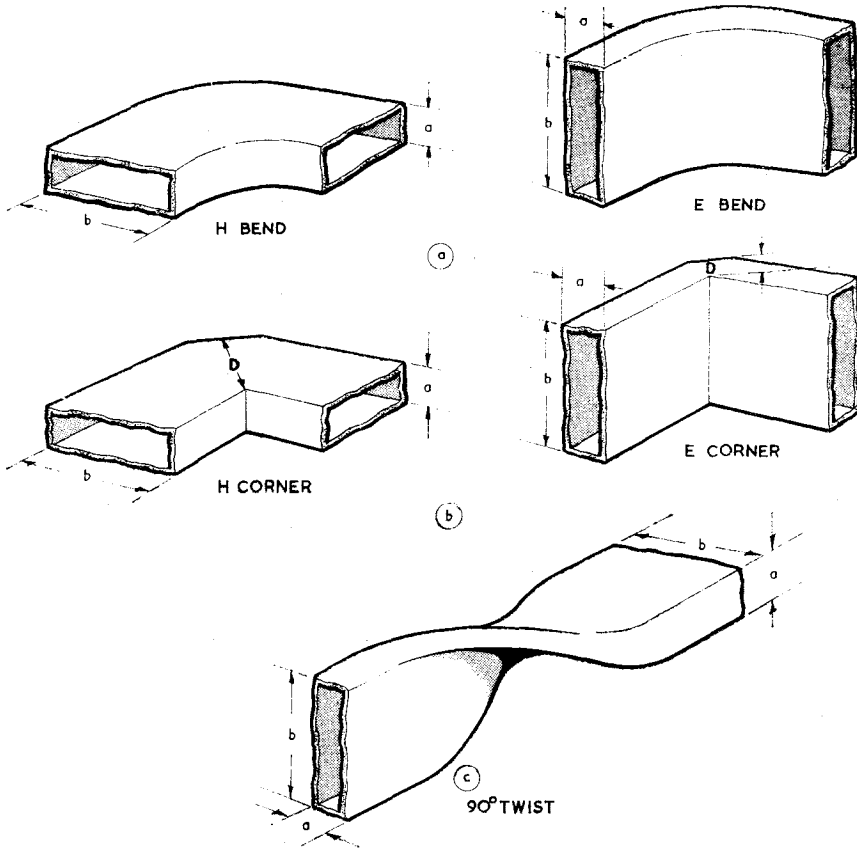


FIG. 19 BENDS, CORNERS AND TWISTS

reflection but this may be kept to a minimum with a gradual bend and by suitable choice of the dimension ' D ' for the corners.

If we wish to turn the E field through 90° , thus changing the plane of polarisation, a gradual twist may be used (Fig. 19*c*).

Flexible Waveguide

Vibrations from a rotating aerial can be transmitted along a rigid waveguide and may damage components mounted on the guide. By inserting a length of *flexible* waveguide close to

the aerial the vibrations are damped. The construction of one form of flexible guide is shown in Fig. 20. It consists of a series of choke discs each with a rectangular hole of waveguide dimensions. The flexible retaining sheath allows movement between the discs while the choke couplings provide electrical continuity of the walls.

Directional Couplers

In many radar systems it is necessary to sample energy coming in one direction only along the guide, i.e. it may be necessary to sample either incident or reflected energy. This can be done quite simply by using a *directional coupler*.

The principle of a two-hole directional coupler is shown in Fig. 21. Waveguide AB is coupled to waveguide CD by two holes spaced $\frac{1}{4}\lambda_g$ apart in the common broad wall. Most of the energy moving from left to right will pass direct to B but some will pass to D by two paths of equal length and will therefore arrive in phase and will total. In the direction of C, however, the two path lengths differ by $\frac{1}{2}\lambda_g$ and therefore the energy will arrive in anti-phase and will cancel.

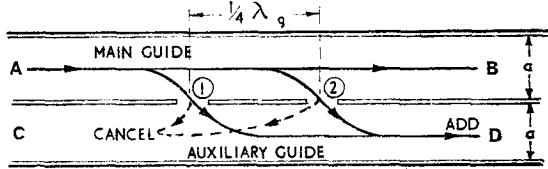


FIG. 21 TWO-HOLE DIRECTIONAL COUPLER

this way the direction of propagation in the main guide governs that in the auxiliary guide. The proportion of energy coupled into the auxiliary guide depends upon the size and position of the coupling holes.

There are many types of directional coupler with two or more holes or slots in either the broad or narrow walls; by increasing the number of coupling holes the bandwidth of the coupler is increased. The guides may be parallel as in Fig. 21 or at an angle.

Waveguide Measurements

A complete waveguide system consists of many components and if any of these break down reflections may occur and standing waves are set up in the guide. As well as reducing the power-carrying capacity and efficiency of the guide, standing waves may reflect a reactance into a microwave oscillator (magnetron, klystron, etc.) coupled to the guide. This can affect the frequency and power output of the oscillator. Standing wave measurements are thus important not only in the design of a waveguide system but also as an indication of faults which may occur in an operational equipment.

In waveguides the voltage standing wave ratio may be defined as the ratio of maximum transverse E field to minimum transverse E field and a waveguide standing wave indicator is designed to measure these two values. A typical example of this instrument is shown in Fig. 22. A slot cut down the centre of the broad side of the guide does not appreciably affect the field pattern and a probe may be inserted in the slot. The voltage induced in the probe is proportional to the amplitude of the voltage standing wave at that point in the guide. This voltage is rectified and its value is shown on a meter. By moving the probe along the slot and noting the maximum and minimum readings obtained the v.s.w.r. can be calculated.

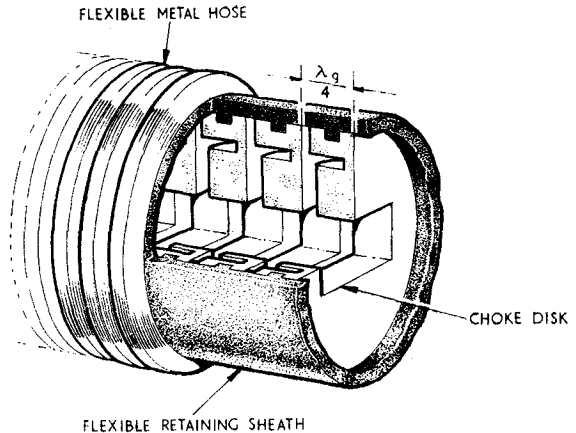


FIG. 20 FLEXIBLE WAVEGUIDE

Similarly, energy moving from right to left will total at C and cancel at D. In

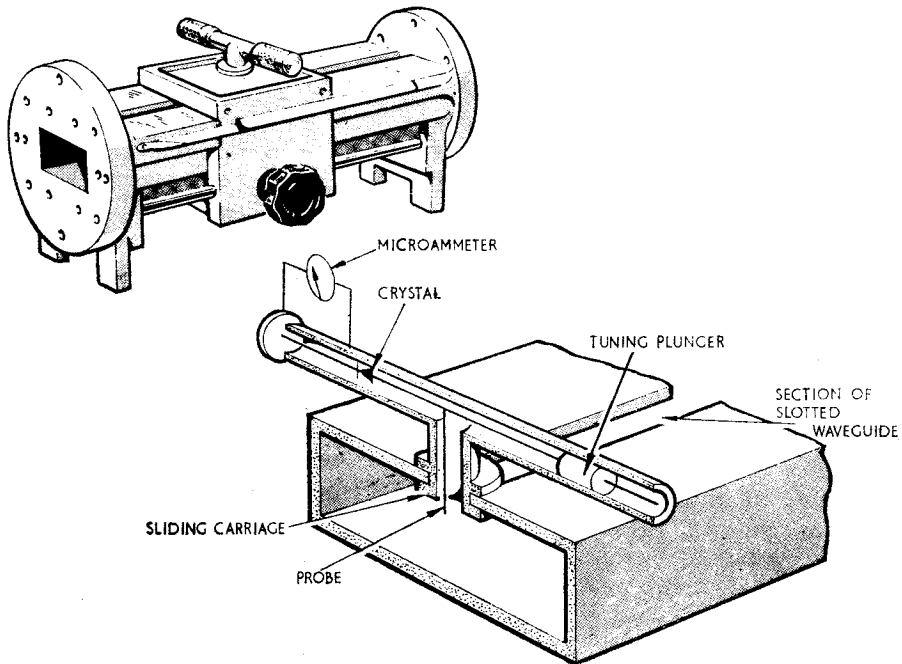


FIG. 22 WAVEGUIDE STANDING WAVE RATIO INDICATOR

Another method of measuring the s.w.r. in a waveguide involves the use of a directional coupler such as the one in Fig. 21. If the transmitter is connected to A and the aerial to B, a matched crystal detector at D will give an output proportional to the incident energy while the output from a detector at C will indicate the reflected energy, i.e. the ratio of the detector outputs gives an indication of the s.w.r.

A s.w.r. indicator of the type shown in Fig. 22 may be used to determine the *frequency* of the energy in the guide by measuring the distance between two successive minima of a standing wave. This measurement is half a guide wavelength and from it the frequency can be calculated.

A more usual method of frequency measurement uses a resonant cavity wavemeter as described in Chapter 1. An accurately calibrated wavemeter may be coupled to an auxiliary guide fed with energy from the main guide via a directional coupler (Fig. 23). Thus standing

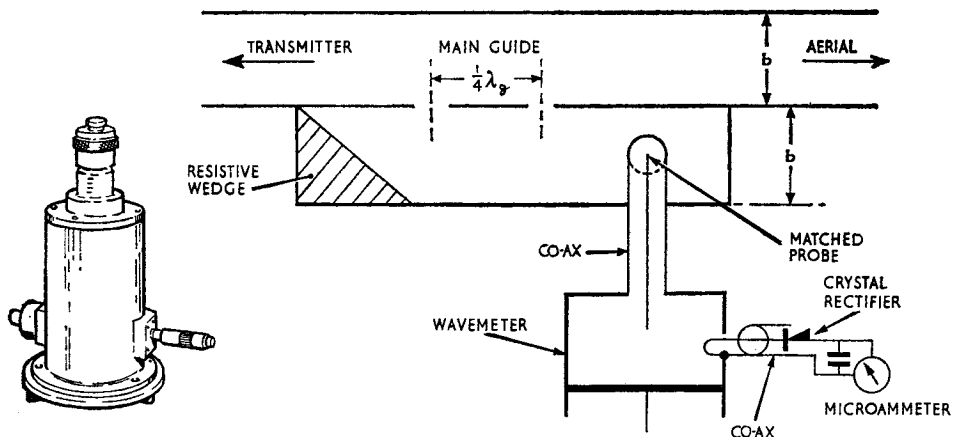


FIG. 23 WAVEGUIDE WAVEMETER

waves in the main guide do not affect the wave-meter which may therefore be left connected to the system to indicate frequency drift. The cavity output is taken to a crystal rectifier and resonance indicator.

The power in a waveguide may be measured by sampling the power via a directional coupler and mounting a temperature-sensitive device such as a *thermistor* in the auxiliary guide (Fig. 24). Absorption of power makes the thermistor hot and a meter measures its change of resistance and hence gives an indication of power in the main guide.

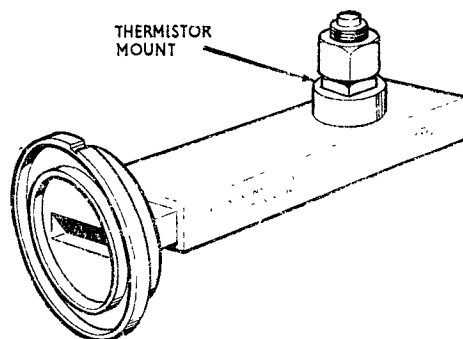


FIG. 24 WAVEGUIDE POWER MEASUREMENT

Waveguide Junctions

In a complex waveguide system it is often necessary to divide the main guide into two branches. This is done with a T-junction, the two forms of which are shown in Fig. 25. The broken lines indicate the equivalent transmission line parallel and series junctions.

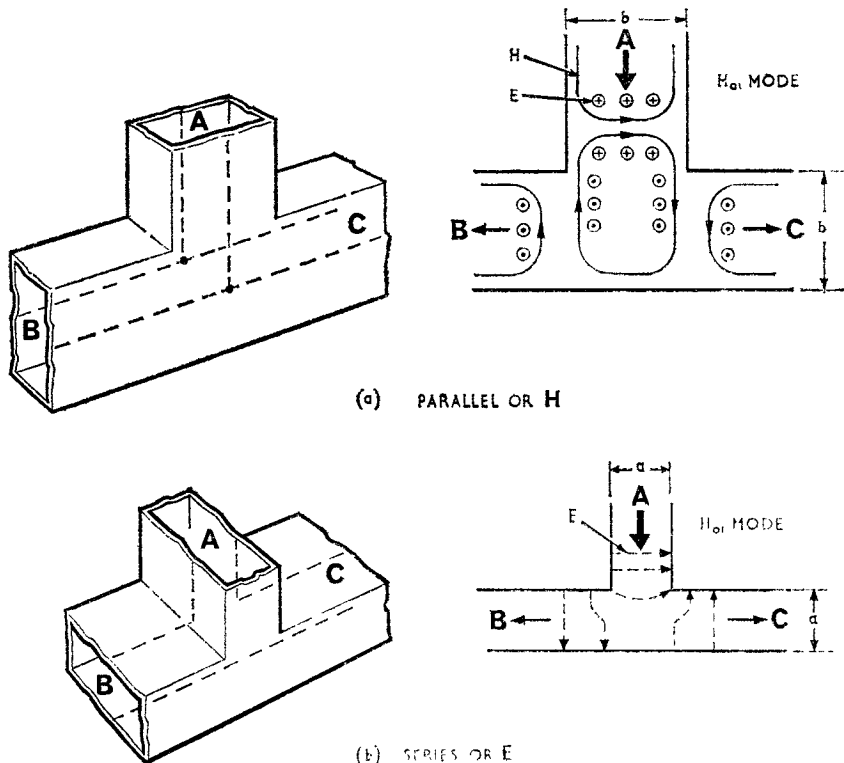


FIG. 25 T-JUNCTIONS

Energy fed into arm A of the *parallel* junction divides equally and with the same E field phase into arms B and C (Fig. 25a). Similarly, if equal strength signals with in-phase E fields are fed into arms B and C they will combine at the junction and provide an output at A. If equal inputs

into arms B and C have anti-phase E fields they will cancel at the junction and no energy will pass through arm A.

In the *series* junction energy fed into arm A divides equally but the E fields in arms B and C are in *anti-phase* (Fig. 25b). If equal strength signals with anti-phase E fields are fed into arms B and C they will provide a combined output at arm A. If the E fields are in-phase, however, there will be no output from arm A.

The junctions introduce discontinuities into the guide but they can be matched by suitably placed irises or matching steps.

The combination of series and parallel arms shown in Fig. 26a is known as the *hybrid T-junction* (sometimes called a "magic T-junction"). It is a very useful waveguide device and forms a part of components such as balanced mixer bridges, aerial switching systems and ferrite circulators (see later).

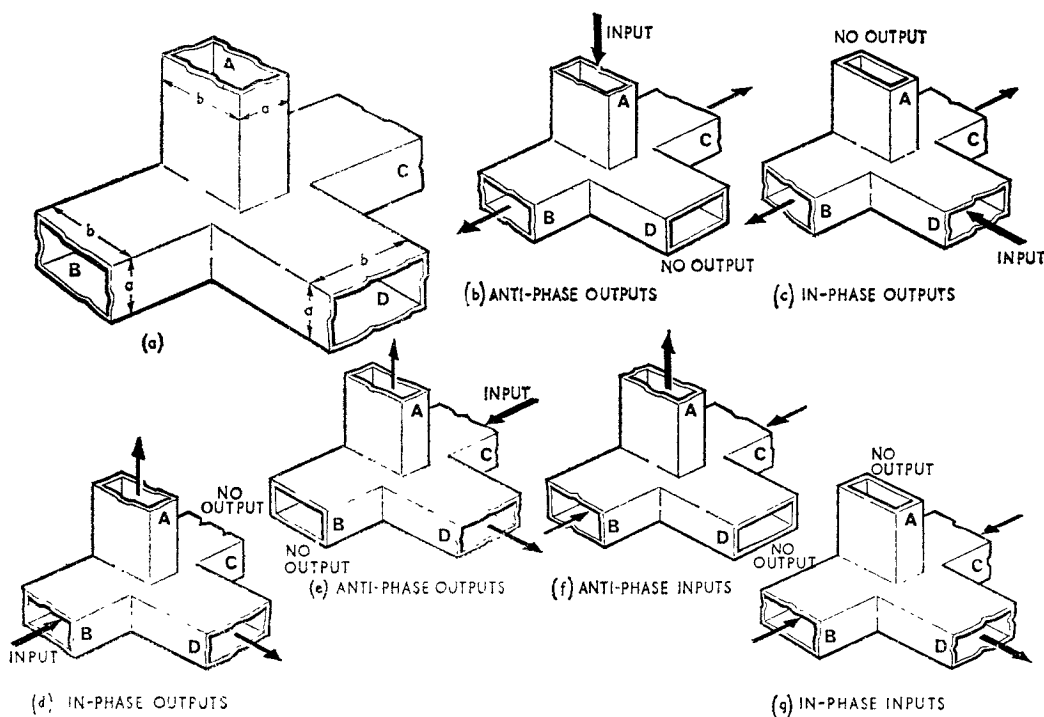


FIG. 26 HYBRID T-JUNCTION

Energy fed into the series arm A divides equally into arms B and C but with anti-phase E fields (Fig. 26b). Energy cannot pass into arm D since the fields are in the wrong planes.

Energy fed into the parallel arm D feeds equally into arms B and C but with in-phase E fields, and cannot feed to arm A (Fig. 26c). Provided the junction is correctly matched an input at B feeds both A and D but not C, and vice versa for an input at C (Fig. 26d and e). Two equal inputs to B and C feed energy to A if the E fields are anti-phase and to D if the E fields are in-phase (Fig. 26f and g).

An alternative to the hybrid T-junction is the *hybrid ring* (or 'rat-race') device shown in Fig. 27. It can be used in a balanced mixer circuit and may be constructed of coaxial cable for use at frequencies below 1Gc/s. Series arms as shown are normally employed.

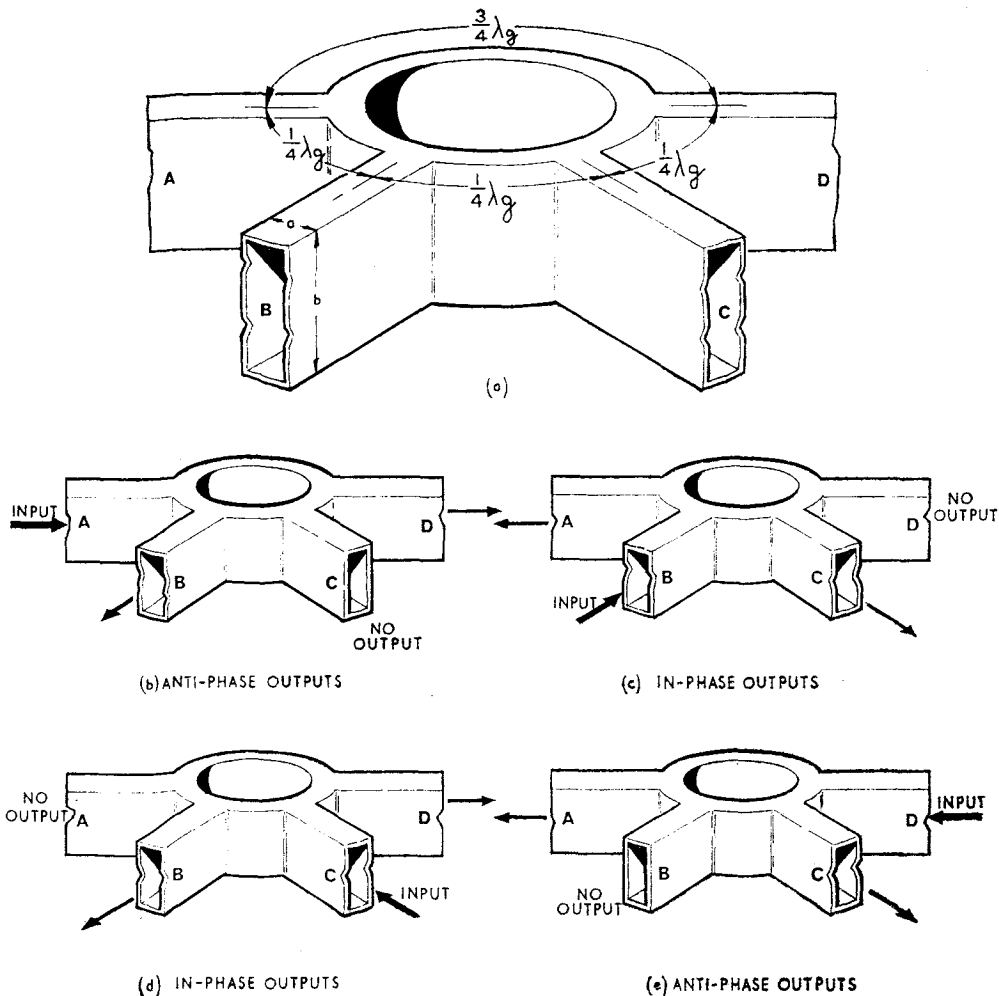


FIG. 27 HYBRID RING

Energy fed into arm A divides equally in the ring and since the two path lengths are equidistant to arm D this arm provides an output. The two path lengths from A to C differ by $\frac{1}{2}\lambda_g$ and so no energy passes through this arm. From A to B the paths differ by λ_g and therefore an output of opposite phase to that from arm D is available at arm B (Fig. 27b).

With an input to arm B the path lengths to both A and C differ by λ_g and therefore in-phase outputs are available from these two arms: the path lengths to D differ by $\frac{1}{2}\lambda_g$ and therefore arm D provides no output (Fig. 27c).

By similar reasoning, an input at C provides in-phase outputs at B and D and no output from A (Fig. 27d); an input at D provides anti-phase outputs at A and C and no output from B (Fig. 27e).

It will be noted that the hybrid ring has the same properties as the hybrid T-junction and for many applications they are interchangeable.

Circular Polarisation

Some radar equipments are required to radiate a circularly polarised wave. This is a wave in which the plane of the E field (and that of the H field) continuously changes from vertical to

horizontal to vertical in the opposite sense and so forth. The fields may rotate in a left-hand or right-hand direction when viewed along the direction of propagation (Fig. 28). Another way of describing a circularly polarised wave is to say that it consists of two plane waves propagating in the same direction but with their E fields at right angles to and 90° out of phase with each other.

A circularly polarised wave may be generated from a plane polarised input by the turnstile junction shown in Fig. 29. When the junction is correctly matched half the energy entering arm A goes into the circular arm E and one quarter into each of the short-circuited arms C and D. No energy enters arm B.

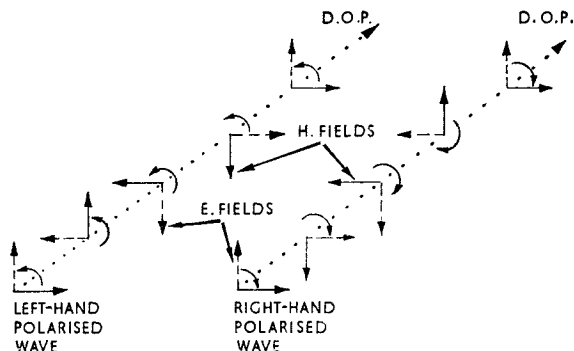


FIG. 28 CIRCULAR POLARISATION

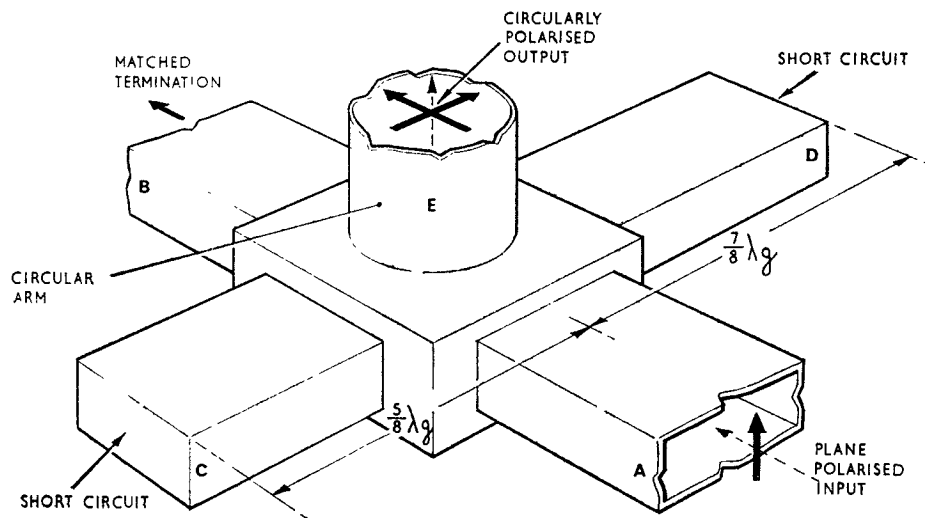


FIG. 29 TURNSTILE JUNCTION

If arm C, measured from the short-circuited end to the centre of the junction is $\frac{5}{8}\lambda_g$ and arm D is $\frac{7}{8}\lambda_g$, the two fields reflected from the ends of these arms combine in phase in arm E and are at right angles to and 90° out of phase with the field entering the arm direct from the input. Thus a circularly polarised wave is available as output from the circular arm. By terminating arm B with a matched load all the power leaking into this arm is absorbed and reflection which would affect the fields in the other arms is avoided. Left-hand or right-hand circular polarisation can be obtained by adjusting the lengths of the short-circuited arms C and D.

When the turnstile junction is used in a radar equipment to generate right-hand polarisation the aerial is connected to the circular arm E and the transmitter and receiver are connected via TR switches to arm A. A *right-hand* circularly polarised echo signal received by the aerial will be converted by the turnstile junction into a plane polarised wave and directed via arm A to the receiver. A *left-hand* polarised echo signal will be directed to the dummy load in arm B. Thus the radar can differentiate between a right-hand and a left-hand circularly polarised wave.

Waveguide Ferrite Devices

Ferrite is an artificial magnetic material with a high resistivity and obtains its magnetic properties due to its free-spinning electrons. When the ferrite is placed in a d.c. magnetic field the axis of spin of the electrons lines up with the direction of the magnetic field. If the spinning electrons are momentarily displaced from alignment by the H component of a microwave field, they will not return directly to alignment but will *precess* about the H field direction rather like a spinning top (Fig. 30). Features of this *gyromagnetic* behaviour are:

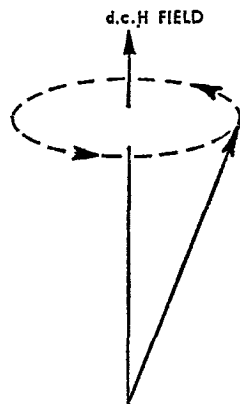


FIG. 30 PRECESSION

a. The *frequency* of precession is proportional to the *magnitude* of the d.c. H field.

b. The *direction* of precession depends upon the direction of the d.c. H field and is clockwise when viewed along the H field direction.

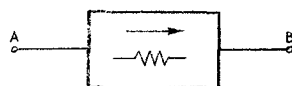
The effect that a magnetised ferrite has on a microwave signal depends upon how close the signal frequency is to the precession frequency; this can be controlled by the d.c. H field magnitude. By varying the strength of the d.c. H field, ferrite materials mounted in waveguides can be used as variable phase shifters and attenuators.

When the magnetising field is adjusted to make the precession frequency equal to the microwave field frequency the ferrite is said to be at *resonance*. Depending upon the direction of electron spin and the position of the ferrite in the guide, a wave in one direction passes through the ferrite with little attenuation but in the opposite direction it is greatly attenuated and the wave energy is converted into heat in the ferrite.

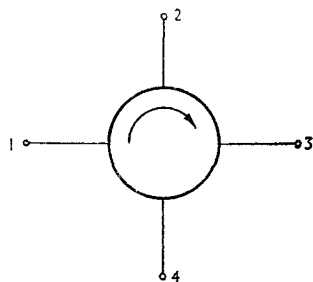
By suitably mounting ferrites in a waveguide and applying a magnetising field of correct magnitude and direction the following non-reciprocal waveguide components can be made:



(a) GYRATOR



(b) ISOLATOR



(c) CIRCULATOR

FIG. 31 WAVEGUIDE FERRITE DEVICES

a. *Gyrator*. This is a two-port device (i.e. it has two waveguide openings A and B in Fig. 31a) which transmits power without attenuation in both directions but imposes a phase difference between the two directions.

b. *Isolator*. A two-port device which greatly attenuates energy in one direction but has little effect on energy in the opposite direction (Fig. 31b).

c. *Circulator*. A multi-port device providing sequential transmission between ports, i.e. energy entering port 1 can leave by port 2 only; that entering port 2 can leave by port 3 only; and so forth (Fig. 31c).

Fig. 32 shows one form of ferrite *gyrator* which imposes a phase change of 180° on energy flowing in direction AB and zero phase change in direction BA.

A rectangular waveguide propagating the H_{01} mode at A is converted via a 90° twist into a circular guide propagating the H_{11} mode and thence into a rectangular H_{01} again. A ferrite rod centrally mounted in the circular section of guide has an axial d.c. magnetising field provided by either a strong permanent magnet or an electromagnet. The ferrite introduces a 90° clockwise rotation of a wave travelling from A to B and this with the 90° clockwise twist gives 180° change at B.

Energy travelling from B to A undergoes a 90° anticlockwise change in the ferrite and a 90° clockwise rotation due to the twist.

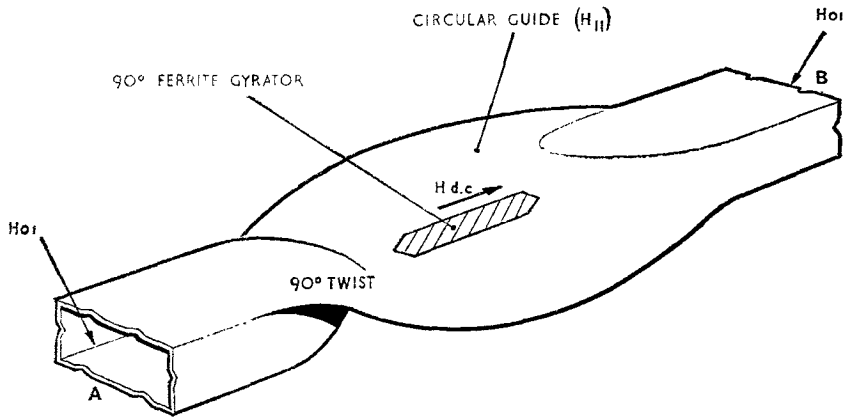


FIG. 32 180° FERRITE GYRATOR

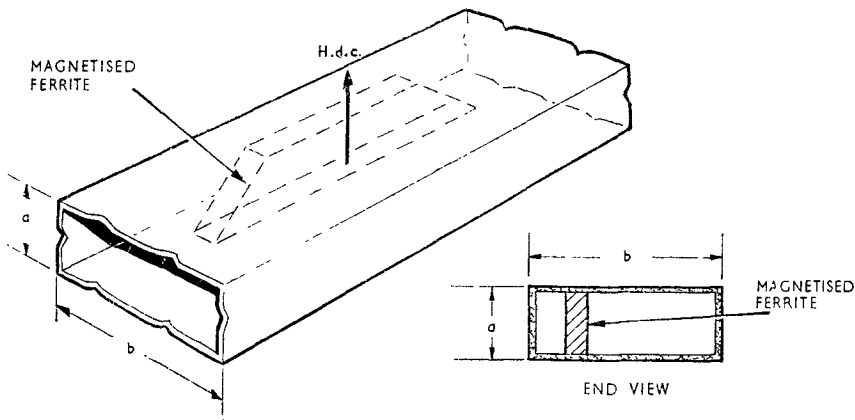


FIG. 33 FERRITE ISOLATOR

Energy at A has therefore the same phase as that injected at B.

A tapered slab of magnetised ferrite mounted in a rectangular waveguide as shown in Fig. 33 acts as a *resonance isolator*. The ferrite is positioned close to the narrow wall; a d.c. magnetic field of the direction shown is set to the resonance value. Heavy attenuation occurs in one direction while in the other direction the attenuation is small. If the direction of the steady magnetising field is reversed the electron spin in the ferrite is reversed and attenuation of a wave travelling in the opposite direction occurs. This also happens if the ferrite slab is positioned on the other side of the guide.

Used in this way a ferrite isolator can absorb energy reflected from a mis-matched load thus improving the s.w.r. while affecting the incident energy only slightly. It also improves frequency stability since variations in the load impedance are isolated from the oscillator. Isolators may also be used between a klystron local oscillator and the mixer to prevent frequency pulling.

A *circulator* may be constructed using a 180° gyrator and two hybrid T-junctions. The principle is shown in Fig. 34.

In the circulator we want power entering by port 1 to leave via port 2 and power entering port 2 to leave via port 3. Similarly, power entering at 3 leaves by 4 and that entering at 4 leaves by 1

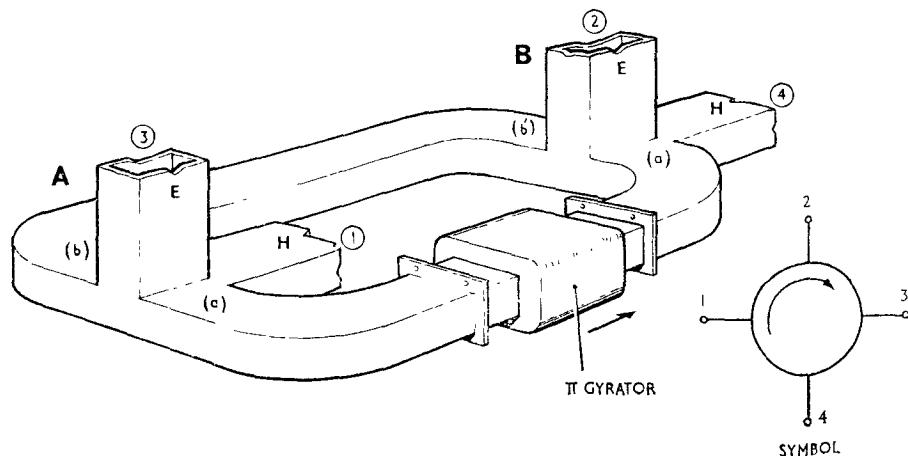


FIG. 34 FOUR-PORT CIRCULATOR

Thus in Fig. 34 power entering junction A via the parallel H arm (port 1) divides equally and in-phase into arms *a* and *b*. No power leaves via the series E arm (port 3). The gyrator shifts the phase of the power in arm *a* by 180° and so the two powers arrive at junction B in anti-phase and therefore leave via the series E arm (port 2).

Power entering the series E arm of junction B (port 2) splits equally but in anti-phase into arms *a* and *b* and no power enters the parallel H arm (port 4). The gyrator in arm *a* does not affect the phase of the power in this arm since propagation is in the reverse direction. Thus the two powers arrive at junction A still in anti-phase and therefore leave via the series arm (port 3).

Circulators can be constructed in various configurations and may be used as aerial switching devices to prevent the large transmitted power damaging the sensitive receiver; e.g. if a transmitter is connected to port 1 of Fig. 34, an aerial to port 2 and a receiver to port 3 the receiver is isolated from the transmitted power. Circulators are also used in parametric devices to separate the output from the input. These applications will be considered in later chapters.

CHAPTER 7

TR SWITCHING AND FREQUENCY CHANGING

Introduction

The first stage of most centimetric radar receivers is the frequency changer or mixer stage. Additive mixing using silicon crystal diodes is used; the crystal is usually mounted in or on part of the transmission system called the *mixer bridge*, and the input echo signal together with the local oscillator voltage is fed into the mixer bridge.

Many radars use a common aerial for both transmission and reception; to connect the aerial to the transmitter output during transmission and to the receiver for the reception period, a *transmit-receive* (TR) switch is necessary. This switch must also prevent the high transmitter power entering the mixer bridge where it would burn out the crystal mixer. Methods of TR switching and crystal mixing will be considered in this chapter.

Microwave Crystal Mixers

The function of a microwave mixer stage is to convert the input frequency to a lower intermediate frequency (usually between 30 and 60 Mc/s). Mixing is done in the first stage of a radar receiver because microwaves cannot easily be amplified by conventional valve and transistor amplifiers. Frequency changing is achieved by applying the input signal and local oscillator output in series to a crystal mixer diode mounted in or on the waveguide or coaxial transmission line. So that the crystal can be conveniently mounted it is manufactured in one of the forms shown in Fig. 1. The principle of the silicon crystal diode and its action as an additive mixer have been dealt with in Part 1 of these notes.

A crystal mixer diode mounted in a coaxial line is shown in Fig. 2. The signal input is picked up by a coupling loop from the TR cell cavity. The crystal is fitted into the inner conductor of the line and held in place by a dielectric rf bypass washer. The local oscillator signal is fed into the line on a coupling probe located approximately a quarter wavelength from the signal coupling loop. This provides a high impedance in the direction of the cavity to the local oscillator signal. Connection between the crystal and the inner conductor of the IF line is provided by a spring contact. The half wave rf choke at the IF side of the crystal isolates the IF line from rf signals.

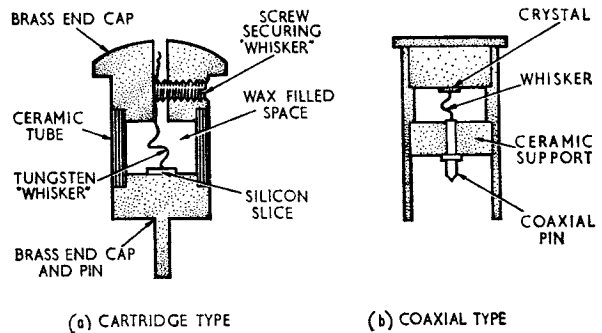


FIG 1. MICROWAVE CRYSTAL MIXER DIODES

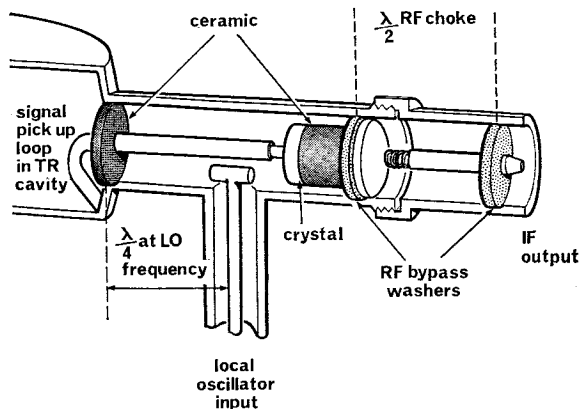


FIG 2. CRYSTAL MIXER IN COAXIAL CABLE

Methods of mounting a single crystal mixer diode for waveguide use are shown in Fig. 3. In Fig. 3a the cartridge, positioned in the centre of the broad dimension, acts as a probe and is matched by a pre-set tuning stub. The fields due to the signal and local oscillator inputs induce voltages in the crystal and mixing takes place. The i.f. output is taken from the mixer via a coaxial feeder to the i.f. amplifiers.

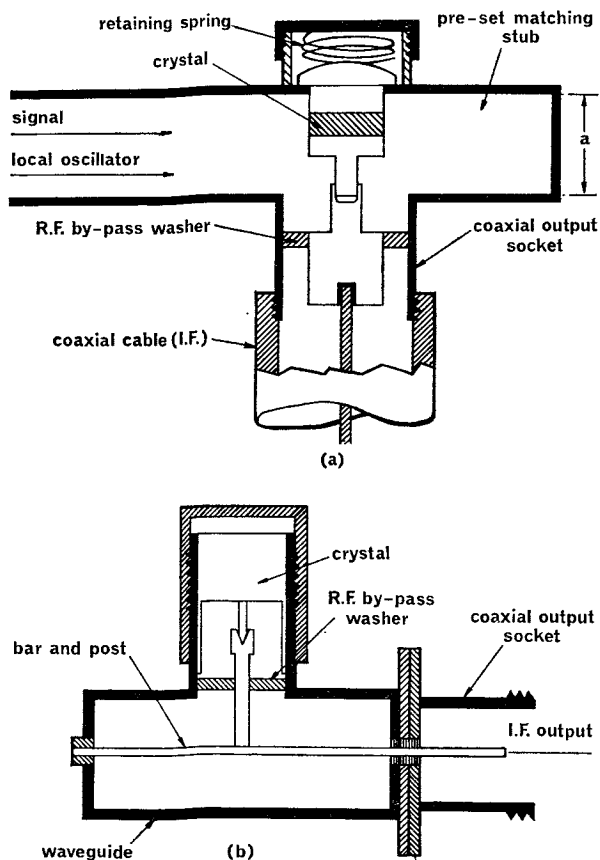


FIG 3. WAVEGUIDE CRYSTAL MIXERS

In the "bar and post" mounting shown in Fig. 3b the crystal is positioned outside the guide and the signal and local oscillator voltages are applied to it via the vertical post. The horizontal bar adds rigidity to the structure and as it is perpendicular to the E field it does not affect the field. This type of crystal mounting is used at X band (3cm) wavelengths and has a fairly wide bandwidth.

Waveguide balanced mixing. In the single crystal mixer the local oscillator, which is usually a low-power klystron, introduces considerable noise. This lowers the signal-to-noise ratio of the receiver thus decreasing the effective range of the radar. By using a hybrid T-junction in a balanced mixer circuit most of the local oscillator noise can be cancelled.

A balanced mixer bridge and its equivalent circuit are shown in Fig. 4. Two crystals are mounted in the closed arms B and C and form a matched load for these arms. The signal from the aerial is fed into the series arm A and the output from the local oscillator into the shunt arm D.

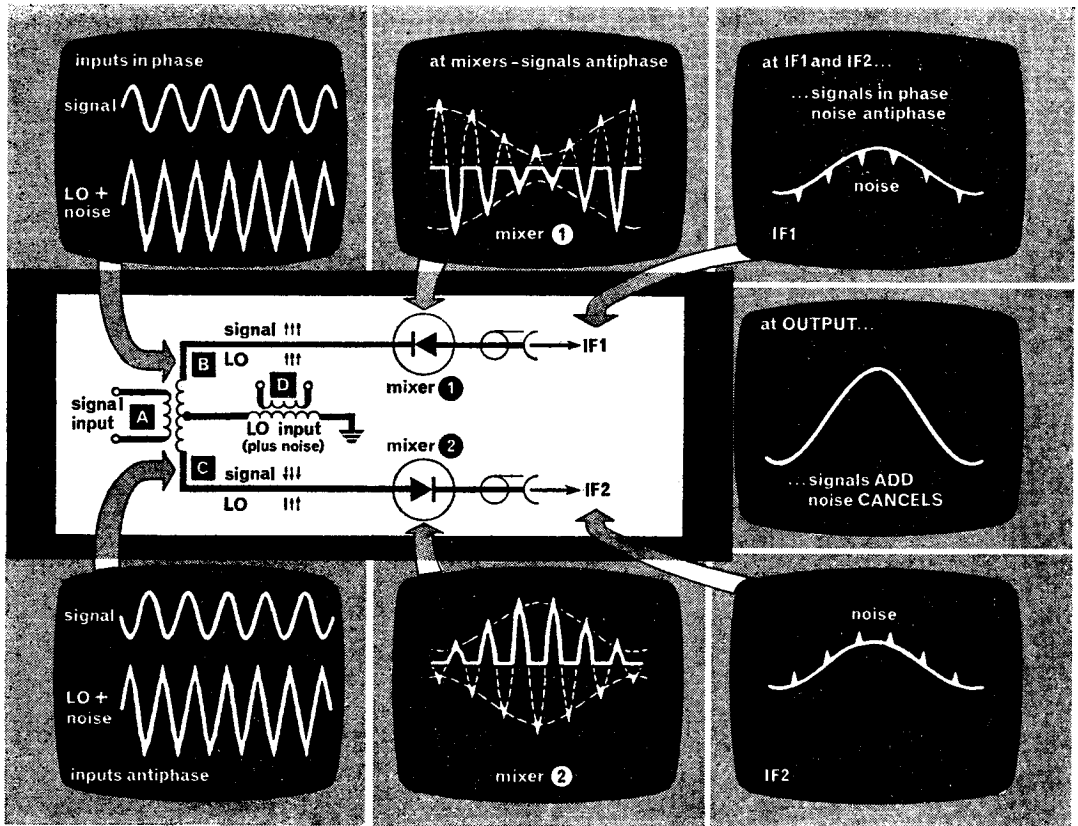
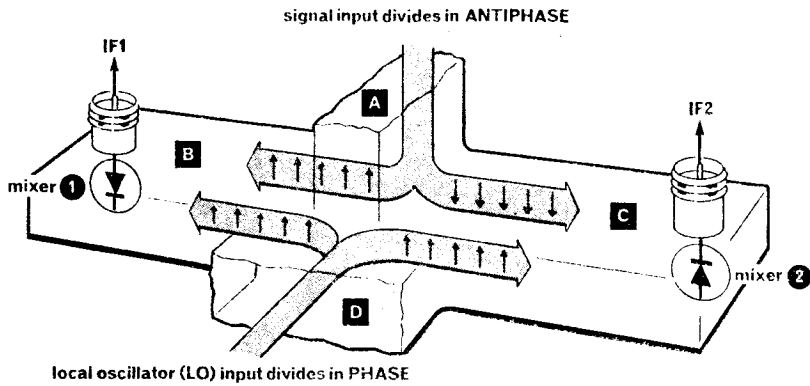


FIG. 4. WAVEGUIDE BALANCED MIXER BRIDGE

The signal applied to arm A divides equally in the two arms B and C, no energy entering arm D; the signals in arms B and C are of equal amplitude but 180° out of phase. The local oscillator input applied to arm D divides into two signals of equal phase and amplitude in arms B and C, no energy entering arm A. At mixer 1 the signal and local oscillator inputs are in phase; at mixer 2 they are 180° out of phase. Thus the i.f. signal at mixer 1 is 180° out of phase with that at mixer 2. However, because the crystals have *opposite polarities*, the signal components of IF1 and IF2 are *in phase* whilst the noise components are in anti-phase. The coupling circuit between the mixers and the i.f. amplifier is such that the in-phase signal components *add* at the mixer output and the anti-phase noise components *cancel*.

The arrangement described is typical of a balanced mixer. It is possible however to have other arrangements. For example, crystals of the *same polarity* may be used if the mixer output is applied to a 'subtractor' coupling circuit, such as a push-pull transformer.

Another important feature of the balanced mixer is the isolation of the aerial signal from the local oscillator circuit and of the local oscillator output from the aerial, both due to inherent properties of the hybrid T-junction.

In addition to the crystal mixer used to produce an i.f. from the signal echo voltage for amplification in the main receiver i.f. amplifier, many radars employ another crystal mixer (often a balanced mixer) to produce a d.c. voltage for *automatic frequency control* (a.f.c.) of the local oscillator output. This section of the mixer bridge is very similar to the signal mixing section but is fed with the local oscillator output and a sample of the transmitter output (in place of the echo signal.) The outline of a microwave mixer bridge employing balanced signal and a.f.c. mixers is shown in Fig. 5.

The silicon crystal diode is widely used as a microwave mixer. It has a low noise factor, does not suffer from transit time effects and has a small input capacitance. Its main disadvantage is its susceptibility to burn-out if high power is applied to it. A crystal loses its sensitivity slowly and should be changed if a radar overall performance check indicates loss of radar performance or if the ratio of reverse resistance to forward resistance falls below 10:1. Even when not in use crystals may be damaged by nearby radiation and so they should be stored in a screened box.

If a mixer crystal burns out shortly after being placed in the transmission system, the transmit-receive device should be checked for serviceability. Crystals are classified according to their frequency range, sensitivity and susceptibility to burn-out.

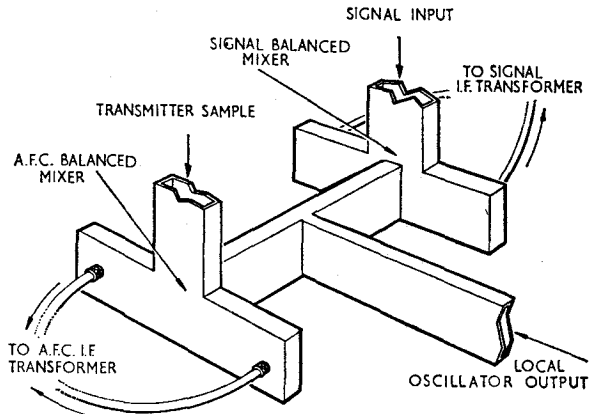


FIG. 5. SIGNAL AND A.F.C. MIXER BRIDGE

Transmit-receive Switching

The transmit-receive switch which connects the common aerial to the transmitter for the duration of the transmission period must provide a high degree of isolation between transmitter and receiver so that the sensitive crystal mixer is not damaged. When the transmitter pulse ends, the switch must rapidly connect the aerial to the receiver input.

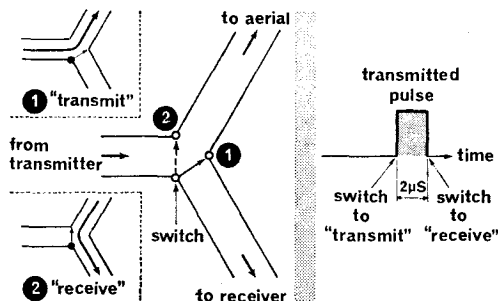


FIG. 6. PRINCIPLE OF TR SWITCHING

coaxial line and operated by a portion of the transmitter r.f. power.

The requirements of such a switch are that it should switch rapidly from receive to transmit allowing a minimum of power to pass to the receiver on transmit and that its "recovery time", i.e. the time it takes to switch the aerial to the receiver, should be very short. An early form of TR switch was a simple spark gap and from this the more efficient *soft rhumbatron* and *multi-cavity TR cells* were developed.

In these devices a resonant cavity is partly enclosed in a gas-filled envelope. On transmit, high-power r.f. enters the cavity, the gas ionises very rapidly and the cavity is thus short-circuited, preventing any further power passing the cell. When the transmitted pulse ends the gas de-ionises very quickly, enabling the echo pulses to pass through the cavity to the receiver.

The TR cell is not a perfect switch; some transmitter power leaks through to the receiver and a small amount of transmitter power is required to maintain ionisation. The recovery time depends upon the type of gas-filling used and on its pressure. Fig. 7 shows a typical voltage waveform across a TR cell; for effective receiver protection the energy in the "spike" must not be too large (the duration of the spike is normally less than 10 nanoseconds) and the de-ionising time must be short (between 1 and 3 μ -seconds).

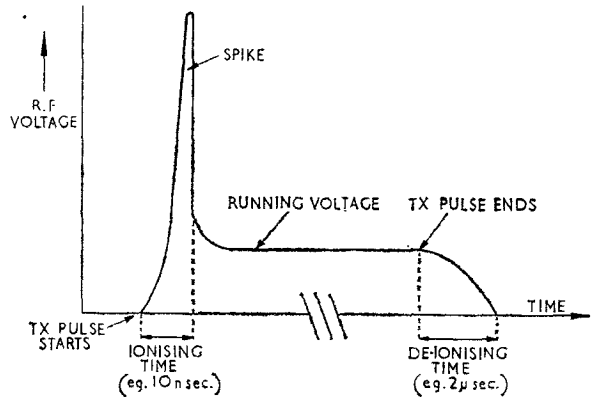


FIG. 7 RF VOLTAGE ACROSS TR CELL

The soft rhumbatron. This TR cell consists of a resonant cavity partly enclosed in a glass envelope filled with water vapour at a low pressure (Fig. 8). The vapour de-ionises rapidly and

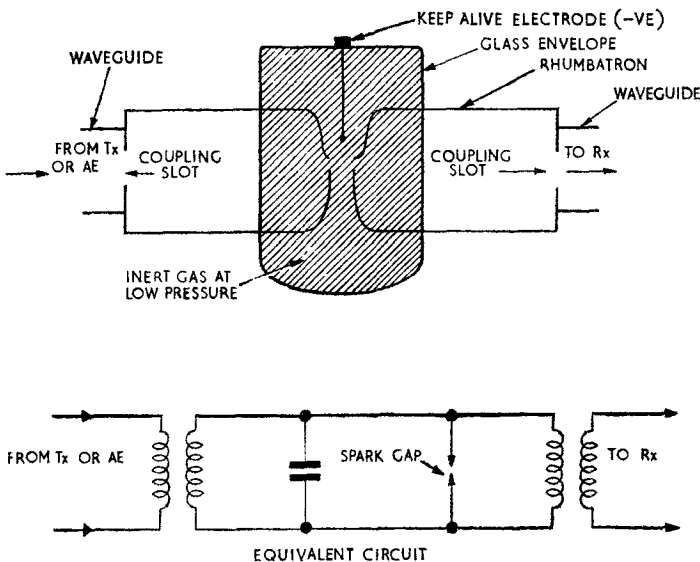
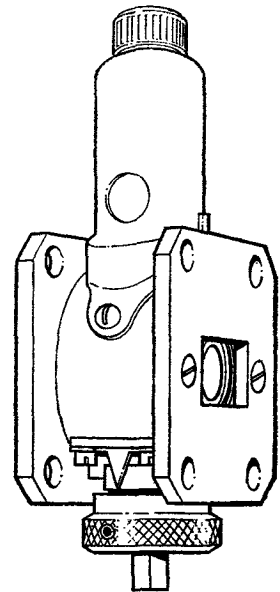


FIG. 8 THE SOFT RHUMBATRON



the soft rhumbatron thus has a rapid recovery time; short-range targets can therefore be detected.

To ensure rapid ionisation of the cell during the initial rise of the transmitter pulse, a "keep alive" electrode is placed close to the cavity lips and maintained at about 1,000 V negative relative to the cavity. A glow discharge results, providing a plentiful supply of electrons near the cavity lips. Thus the gas ionises rapidly and the receiver is adequately protected.

The rhumbatron cavity is coupled to coaxial cable via coupling loops and to waveguide via slots. During reception the rhumbatron acts as a coupling circuit to the receiver. Some soft rhumbatrons may be tuned to resonance mechanically.

Multi-cavity TR cell. This form of TR switch is more efficient than the soft rhumbatron and because it consists of several cavities it has a wider bandwidth. As shown in Fig. 9 it consists of a gas-filled section of waveguide sealed at both ends with glass windows through which r.f. power can pass. Two spark gaps spaced as indicated are mounted in the gas-filled cavities formed

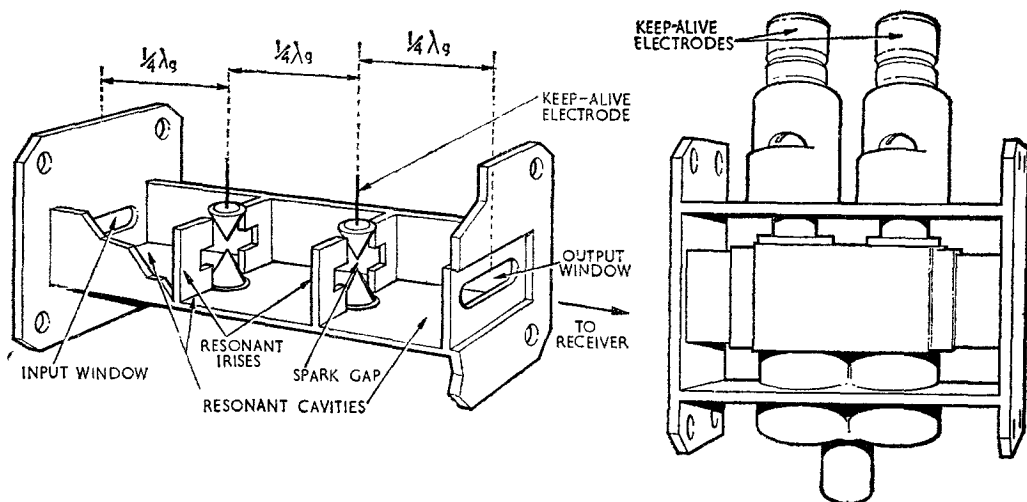


FIG. 9 MULTI-CAVITY TR CELL

by resonant irises. A keep-alive electrode is placed in the spark gap furthest from the input window and when the transmitter fires this gap breaks down first. When it does so a standing wave with a voltage maximum at the first spark gap is set up in the cell. This helps the first spark gap to break down causing a voltage maximum at the input window which then arcs across. Because of the two spark gaps this sequence occurs very rapidly. In some multi-cavity TR cells a keep-alive electrode is provided at each spark-gap (Fig. 9). During reception the cell acts as a stagger-tuned broad-band coupling circuit.

Transmitter-blocking (TB) cell. A simple form of spark gap is often used in conjunction with a TR cell to direct the echo pulse power into the receiver and to block it from the transmitter oscillator. This is known as a *transmitter-blocking* (TB) cell, sometimes called an *anti-TR* (ATR) cell. It may consist of a simple resonant iris mounted inside a gas-filled envelope; or a gas-filled resonant cavity $\frac{1}{4}\lambda_g$ long. Its position in the transmission system is such that it has no effect on the system during transmission and so it does not require a keep-alive electrode.

The keep-alive electrode in a TR cell generates noise, as do all gas-discharge devices. About $\frac{1}{2}$ to 1 db transmitter power is required to maintain the arc during transmission and the cells introduce a loss during reception. These factors reduce the radar performance. In high-power TR cells the transparent windows may overheat and crack. The normal life of a TR cell is a few hundred hours, and the cell must be suspected if several mixer crystals in succession burn out, if the maximum range of the radar falls, or if short-range targets cannot be detected.

Branch-arm TR Switching

The principle of TR switching using TR and TB cells mounted in series arms is illustrated in Fig. 10. A waveguide system with its twin-wire equivalent is shown although the principle applies also to coaxial cable transmission lines.

At the start of the transmitter pulse high-power r.f. flows towards the aerial, enters the TB cell and produces an arc at the spark-gap. This short-circuit is reflected over $\frac{1}{2}\lambda$ as a short-circuit completing the main transmission line. The r.f. power thus moves on towards the aerial.

RF energy enters the TR cell causing the soft rhumbatron to arc across and this short-circuit is reflected over $\frac{1}{2}\lambda$ as a short-circuit completing the main line. Little further power enters the TR cell and the receiver is protected, most of the r.f. power going to the aerial.

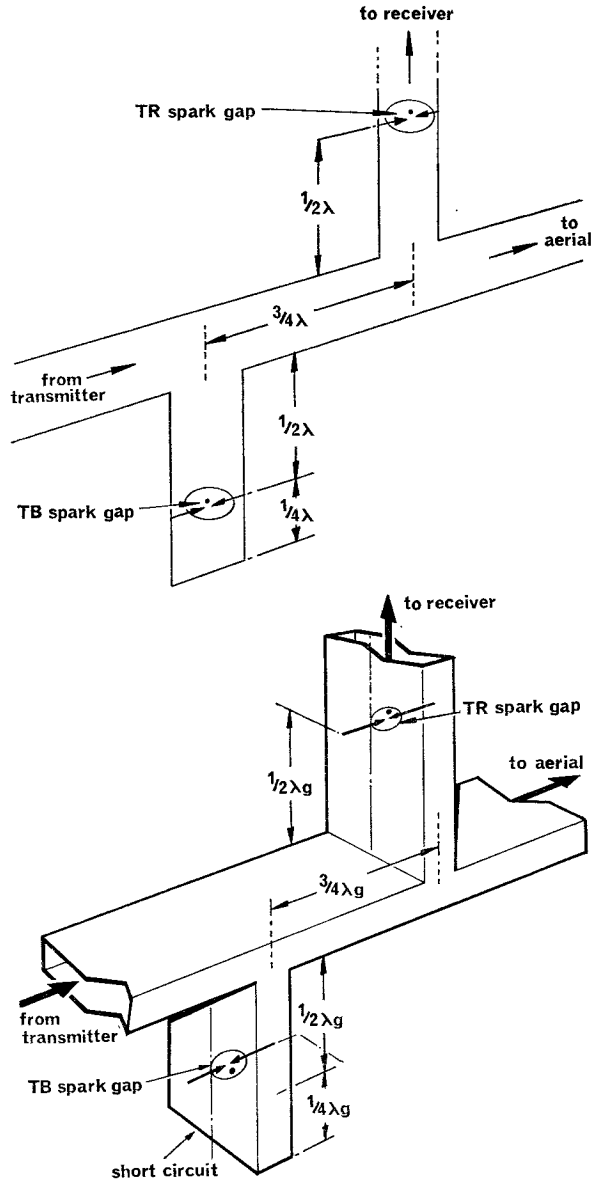


FIG 10. PRINCIPLE OF TR SWITCHING USING SERIES-MOUNTED TR AND TB CELLS

When the transmitted pulse ends both cells rapidly de-ionise and become open-circuits. The received echo power is too weak to ionise the cells and the short-circuit termination in the TB cell is reflected as an open-circuit to the main line. This, over an odd number of quarter wavelengths, is reflected as a short-circuit at the TR cell junction. Thus the echo power is directed through the TR cell to the receiver and is not absorbed in the transmitter.

Either series or shunt arms, or a combination of both may be used in a TR switching system. Fig. 11 shows a system using shunt-mounted TB and TR cells. The distances, measured in multiples of $\frac{1}{4}\lambda$ ($\frac{1}{4}\lambda_g$ for waveguides), are such that the receiver is protected when transmitting and maximum echo power is developed at the receiver during reception. In some TR systems the TB cell is dispensed with by making the distance from the transmitter to the TR cell junction an *even* number of quarter wavelengths. Thus the high output impedance of the transmitter when it is switched off is reflected as a high impedance at the TR junction and little echo power flows towards the transmitter.

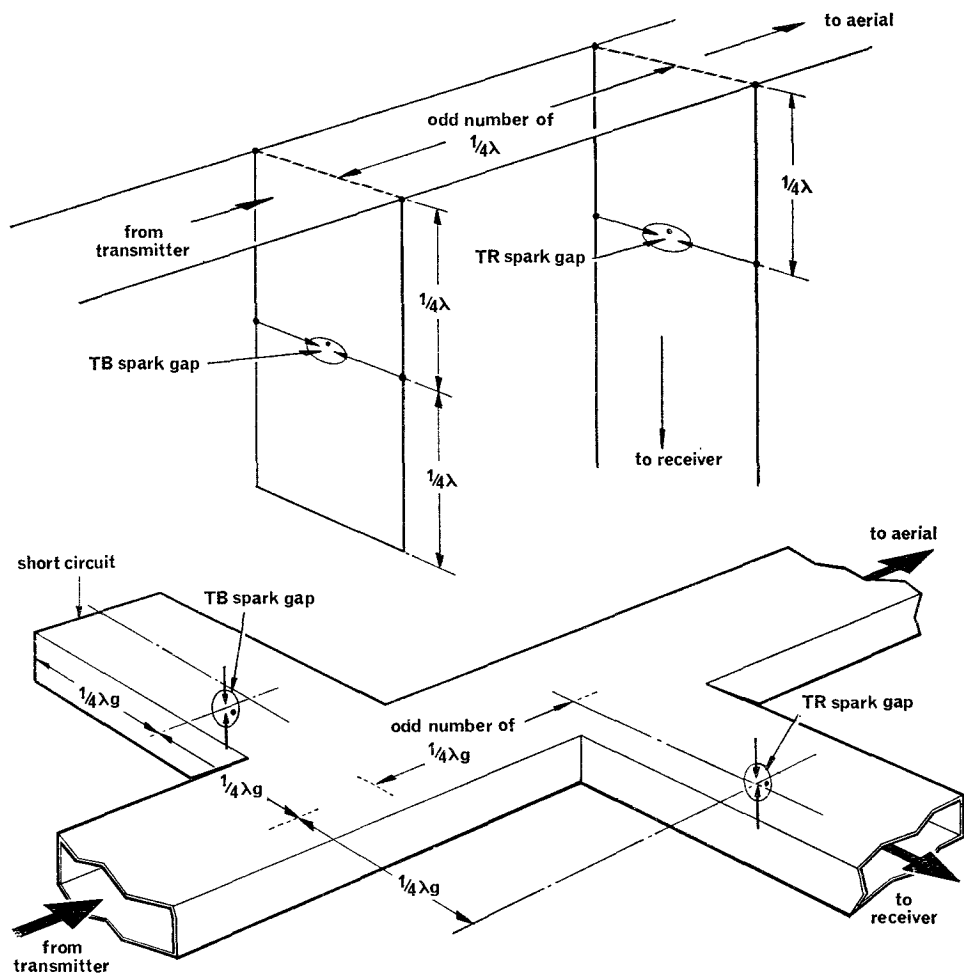


FIG 11. PRINCIPLE OF TR SWITCHING USING SHUNT CELLS

Balanced TR Switching

The branch-arm system of TR switching is narrow-band, i.e. the efficiency and effectiveness of the system is reduced if the transmitter frequency changes. The *short-slot hybrid* system of TR switching has a wider bandwidth and since directional coupling is used in conjunction with

TR cells the receiver isolation is improved.

A short-slot hybrid consists of two sections of waveguide with a common *narrow* wall. Over a short length of waveguides the common wall is removed (Fig. 12a) and power fed in at A divides into the two guides. The slot distorts the field pattern so that half the power fed in at A is coupled towards D and half towards B (Fig. 12b). No power is coupled towards C. Thus the short-slot is a *3db coupler* and has the important property that the power available at D, having passed through the slot, has had its phase *advanced* by 90° relative to that at B. In some types of short-slot 3db coupler the phase at D is *retarded* by 90° relative to that at B.

Fig. 12c shows the power distribution and phase when an input is applied to C.

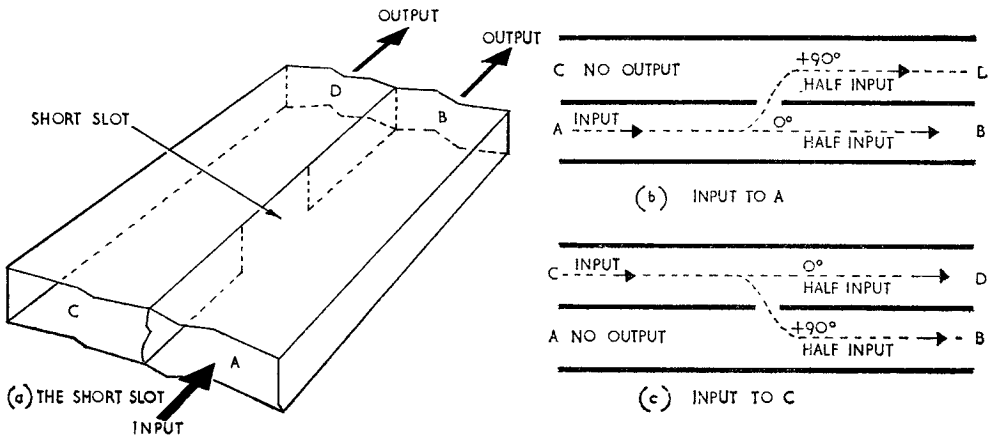


FIG. 12 PROPERTIES OF THE SHORT-SLOT 3db COUPLER

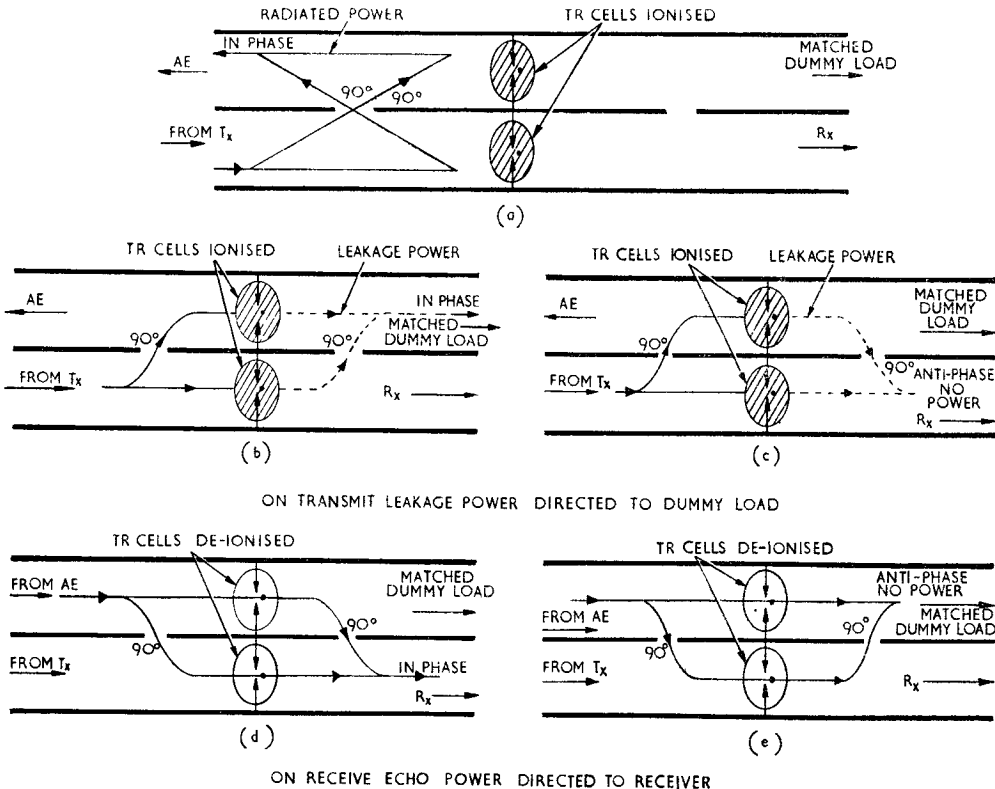


FIG. 13 BALANCED TR SWITCHING USING SHORT-SLOT HYBRIDS

When used as a TR switch two short-slot couplers are arranged as shown in Fig. 13. During transmission the two TR cells ionise and reflect the power as shown by the solid lines in Fig. 13*a*. Since both paths pass once through the slot the power totals in phase towards the aerial. Leakage power through the TR cells (shown by the broken lines in Fig. 13 *b* and *c*) will cancel in the path to the receiver and is directed into the matched dummy load.

During the reception period echo power passes through the de-ionised TR cells as shown in Fig. 13*d* and *e*. Because of the 90° phase change due to the slots the power totals in phase in the receiver arm and cancels in the dummy load arm.

Another form of balanced TR switch using two hybrid T-junctions is shown in Fig. 14.

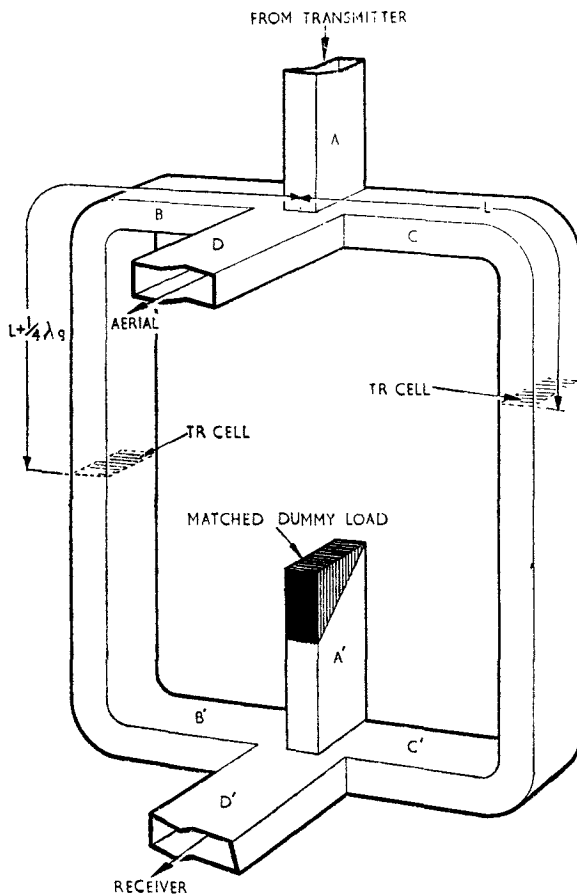


FIG. 14 BALANCED TR SWITCHING USING HYBRID T-JUNCTIONS

During transmission, power from arm A divides equally into arms B and C and the two TR cells ionise, placing a short-circuit in these arms. The E fields entering arms B and C are in anti-phase and are reflected at the TR cells back towards the junction. In arm B the path length from the junction to the TR cell and back is $\frac{1}{2}\lambda_g$ longer than that in arm C, so the two fields arrive back at the junction in phase and therefore leave by arm D to the aerial.

In addition to the protection afforded by the TR cells the lower hybrid T-junction provides further receiver isolation. The power which leaks through the TR cells is in anti-phase and since the path lengths BB' and CC' are equal this leakage power will pass into the dummy load in arm A' and none will enter the receiver arm D' .

During the reception period the echo power from the aerial in arm D divides equally and in phase into arms B and C . It passes through the de-ionised TR cells and combines, still in phase, to leave by D' to the receiver. No echo power enters the dummy load or the transmitter.

Waveguide shutters. When the radar transmitter and receiver are switched off there is no e.h.t. voltage on the keep-alive electrode of the TR cell. Thus considerably increased power is required before the cell ionises and provides protection for the receiver crystal. High-power radiation from a nearby transmitter operating on the same frequency may therefore damage the receiver crystal. To prevent this, a solenoid-operated metal shutter which short-circuits the receiver input is sometimes provided.

Ferrite TR Devices

The directional properties of hybrid junctions can be used in conjunction with ferrite phase shifters to provide TR switching devices without TR or TB cells. The four-port circulator discussed on page 303 is one example and is shown again in Fig. 15.

The transmitter is connected to the H arm of the hybrid T-junction A (port 1) and the receiver to the E arm (port 3); thus the receiver is isolated from the transmitter power. The transmitted power divides equally and in phase into arms a and b , but because of the 180° ferrite phase shifter in arm aa' , the power in the two arms arrives at the second junction in anti-phase and therefore leaves by the E arm (port 2) to which the aerial is connected. Leakage power is absorbed

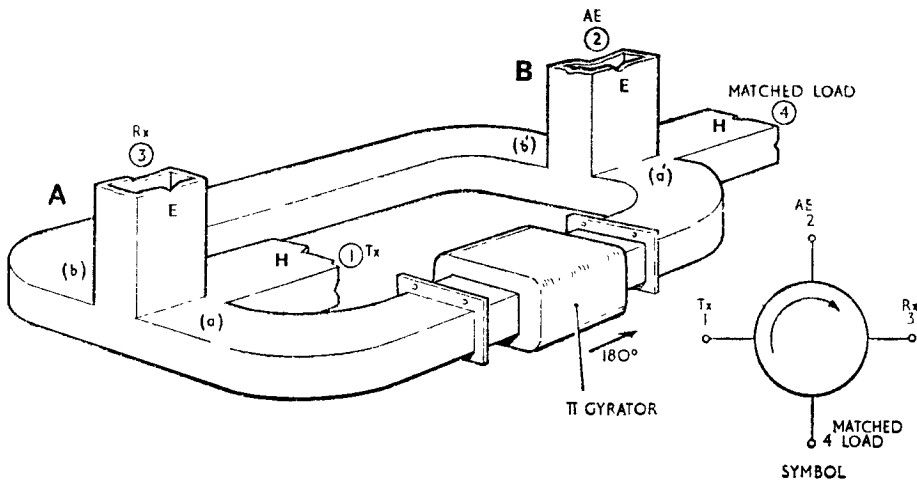


FIG. 15 FERRITE TR SWITCHING

by the matched load in arm 4. To provide increased receiver isolation a TR cell may be mounted in the receiver arm to protect the receiver against leakage from port 1 to port 3 and against power reflected from the aerial.

The echo power enters by port 2 and divides equally and in anti-phase into arms a' and b' . It does not undergo any phase change in arm a' since the phase-shifter is a non-reciprocal device

and therefore the echo power enters the E arm (port 3) to the receiver.

If the receiver isolation is sufficient without a TR cell, this device may be used in CW radar systems to isolate the receiver from the transmitter. As no switching devices are employed, the receiver is isolated from the transmitter at all times.

A three-port circulator feeding into a switched three-port circulator as in Fig. 16 can be used in a pulsed radar system to give increased receiver protection. Transmitter power from port 1 enters port 2 to the aerial. Leakage power, and power reflected from a mis-matched aerial is directed into the matched dummy load in port 3, none entering the receiver in port 4.

At the end of the transmitter pulse the ferrite d.c. magnetising field of circulator B is reversed and thus echo power from the aerial is fed to the receiver.

Ferrite TR systems are fairly broad-band and have longer life and shorter recovery time than systems using gas-filled TR switches. However, they are usually bulkier and heavier because a permanent magnet or solenoid is required to provide a d.c. magnetic field.

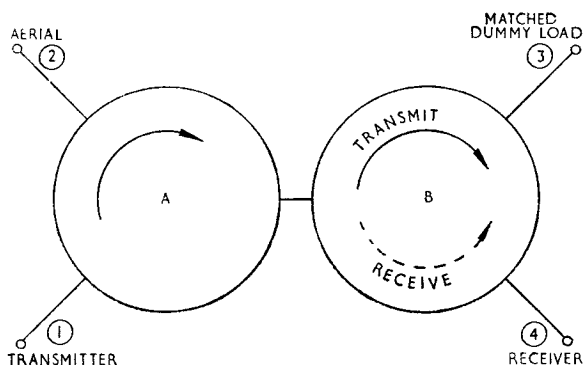


FIG. 16 SWITCHED FERRITE TR SWITCH

CENTIMETRIC AERIALS

Introduction

One advantage of using centimetric wavelengths is that it is possible to build small aerials to give narrow beams. As the wavelength used gets shorter so the size of aerial required to produce a given beamwidth decreases. For example, at 300 Mc/s (100 cm) an aerial approximately 197 feet across is required to produce a beamwidth of about 1° ; at 10,000 Mc/s (3 cm) the same beamwidth is obtained with a 6 foot aerial.

With narrow beams accurate bearings in azimuth and elevation can be obtained and, because of the small size and light weight involved rapid movement of the aerial is possible. Size and weight are important considerations in airborne radars.

We can produce a narrow beam by radiating energy from an aerial into a specially-shaped *reflector* in much the same way as a searchlight beam is formed. At a given wavelength the larger the reflector, the narrower will be the beam. This is the most common type of centimetric aerial and is known as a *reflector aerial* (Fig. 1a).

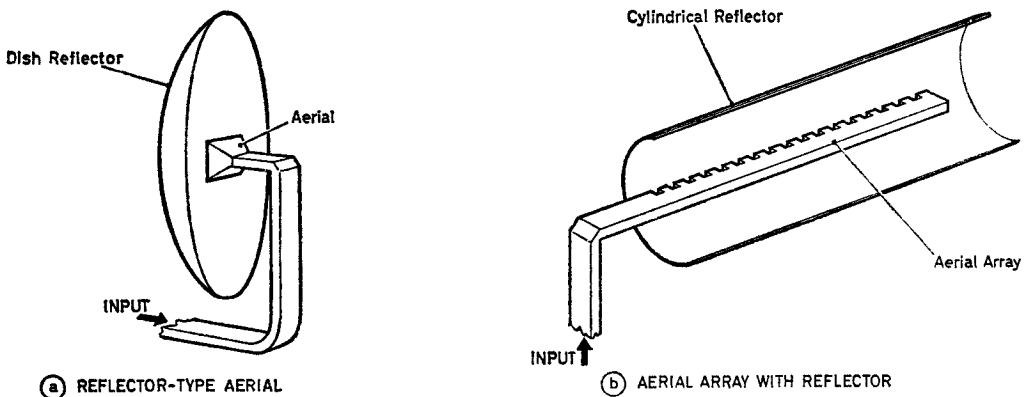


FIG. 1 TYPES OF DIRECTIONAL CENTIMETRIC AERIAL

Another type of centimetric radiator, the *aerial array*, consists of a number of aerials fed in such phase relationship as to form a beam in the required direction. By altering the phase of the energy fed to the aerials the direction of the beam can be altered; by increasing the number of aerials the beam is made narrower. Often an aerial array is used with a reflector to produce a very narrow pencil-shaped beam (Fig. 1b).

In this chapter we shall consider these types of directional aerial. The same general properties apply to both transmitting and receiving aerials.

Aerial Parameters

In order to measure the effectiveness of a directional aerial several parameters are used.

Beamwidth. The width of the radiated beam of a directional aerial is measured by the angle between the two points on the power radiation diagram where the power has fallen to half its

maximum value (i.e. fallen by 3db). Often the radiation diagram is plotted in field strength units and as field strength is proportional to the *square root* of power the corresponding points are where the field strength has fallen to $\sqrt{\frac{1}{2}}$, i.e. 0.707 of the maximum field strength (Fig. 2). On both power and field strength diagrams the beamwidth is measured between the - 3 db points.

Gain. The power gain of a directional aerial is the ratio of power radiated in the direction of maximum radiation to that radiated from a reference aerial (which may be an omni-directional aerial or a simple dipole), both aerials being fed with the same current, i.e.

$$\text{Gain} = \frac{\text{Max. power radiated from directional aerial.}}{\text{Power radiated from reference aerial.}}$$

The gain is usually measured in decibels.

Radiation patterns. We know that by measuring the field strength in a certain plane around an aerial we can plot a polar diagram for the aerial in that plane. Fig 3a shows a polar diagram

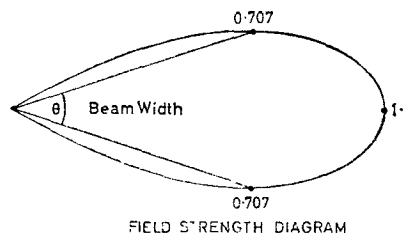
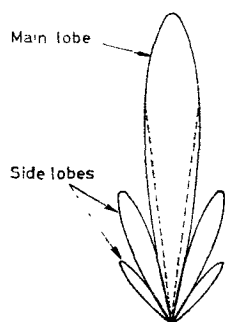
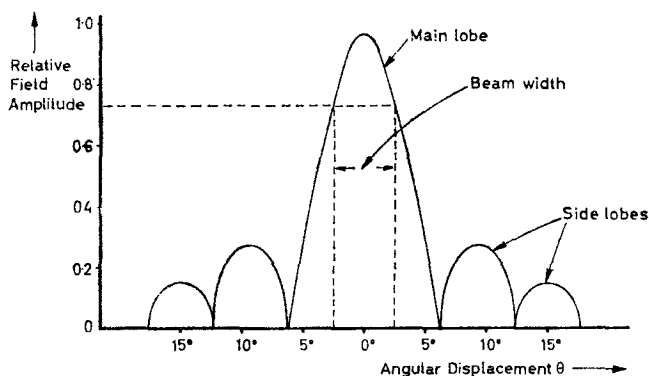


FIG. 2 BEAMWIDTH



(a) POLAR CO-ORDINATES



(b) CARTESIAN CO-ORDINATES

FIG. 3 RADIATION PATTERNS

for a typical centimetric aerial. On a narrow diagram it is difficult to measure the beamwidth and often the plot is made in *Cartesian co-ordinates* as shown in Fig. 3b.

From this radiation diagram we can clearly see that the beamwidth is 5° . Notice that as well as the main lobe there are several *side lobes*. In centimetric radar aerials these must be kept as small as possible, for as well as wasting energy they can cause false bearing indications.

Polarisation. An e.m. wave radiated from an aerial consists of alternating E and H fields at right angles to each other and also at right angles to the direction of propagation (Fig. 4). At some distance from the aerial the plane formed by the E (or H) vector and the direction of propagation is in a fixed direction and the wave is called a *plane wave*. The plane which contains the E vector and the direction of propagation is called the *plane of polarisation* and if this is vertical relative to the earth we have a *vertically polarised wave* (Fig. 4a); if it is horizontal the wave is *horizontally polarised* (Fig. 4b).

Most radars employ horizontal or vertical polarisation but for some purposes (for example, to cancel echoes due to rain) it is necessary to vary continuously the plane of polarisation. This means that the E and H vectors, still at right angles to each other, rotate as the wave progresses. If the r.m.s. values of the E and H vectors remain constant the wave is said to be *circularly*

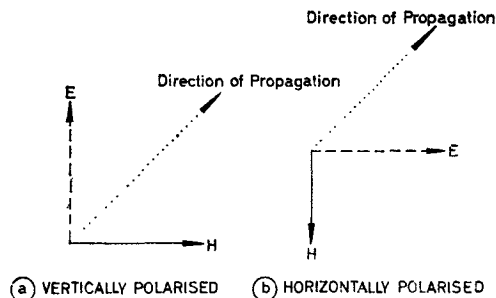


FIG. 4 THE PLANE WAVE

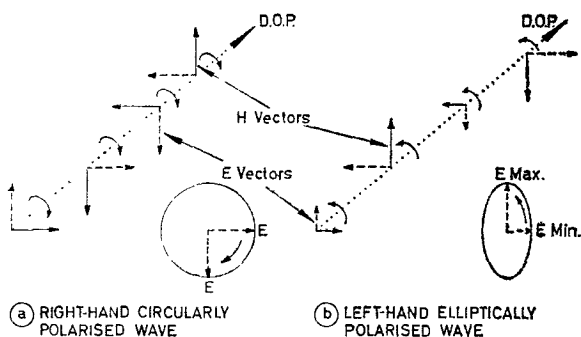


FIG. 5 CIRCULAR AND ELLIPTICAL POLARISATION

polarised (Fig. 5a). If the r.m.s. values vary sinusoidally the wave is *elliptically* polarised (Fig. 5b).

When the E vector rotates in a clockwise direction, viewed in the direction of propagation, the wave is a clockwise or *right-hand* polarised wave as in Fig. 5a; if the E vector rotates in the opposite direction it is said to be an anti-clockwise or *left-hand* polarised wave (Fig. 5b).

Parabolic Reflectors

At centimetric wavelengths a specially-shaped metal reflector, similar in principle to a searchlight mirror, is used to produce a narrow beam. The *parabolic reflector* is the most widely used. In Fig. 6a, ABC is a cross-section of a parabolic reflector; point F is the *focal point*. If

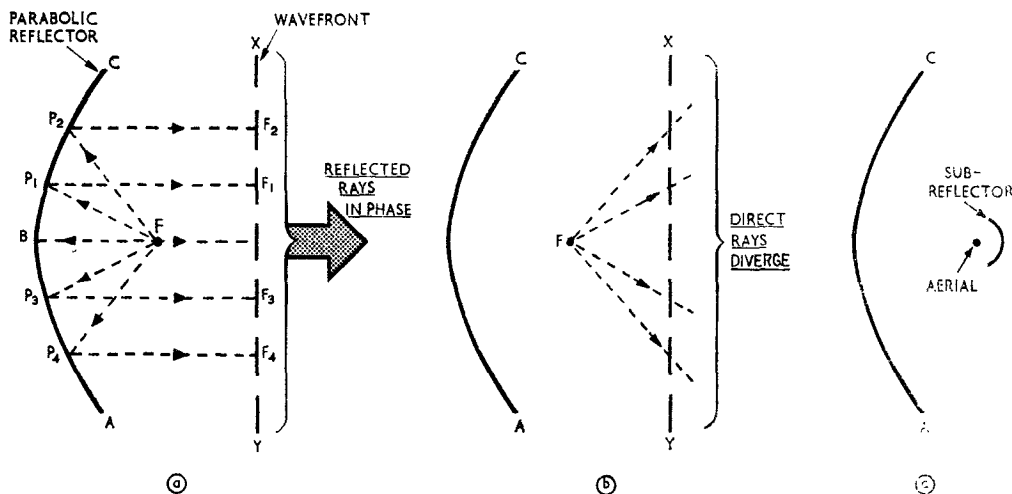


FIG. 6 PRINCIPLE OF THE PARABOLIC REFLECTOR

an aerial is placed at the focal point all energy radiated towards the reflector is reflected parallel to the axis BF. Furthermore, the path lengths FP_1F_1 , FP_2F_2 , etc. are all equal. Thus rays radiated from F towards the paraboloid are reflected as parallel rays travelling in the same direction and arrive at XY in phase and therefore reinforce each other. A narrow beam in the direction of the arrow is thus formed by the reflector.

Energy radiated from F directly towards XY without striking the reflector will not be in phase at XY and will produce diverging rays (Fig. 6b). This effect can be reduced by placing a small shield or sub-reflector behind the aerial as in Fig. 6c.

The larger the area of the opening (or aperture) AC, measured in wavelengths, the narrower will be the beam. The power gain of a parabolic reflector with an aperture diameter D is given approximately by:

$$G = 6 \left(\frac{D^2}{\lambda^2} \right).$$

where λ is the wavelength of the radiated energy. D and λ must be measured in the same units.

Types of parabolic reflector. To produce a pencil-shaped beam of circular cross-section a parabolic reflector (or 'dish') with a circular aperture as shown in Fig. 7 is used.

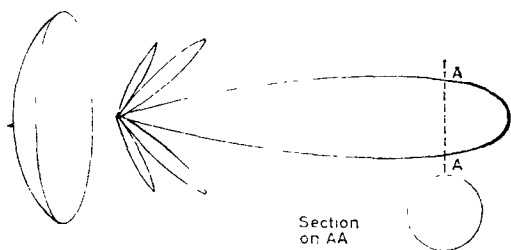
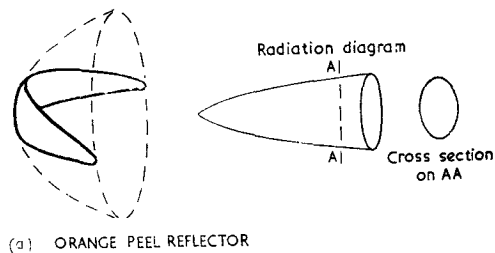


FIG. 7 PARABOLOID OR PARABOLIC DISH



(a) ORANGE PEEL REFLECTOR

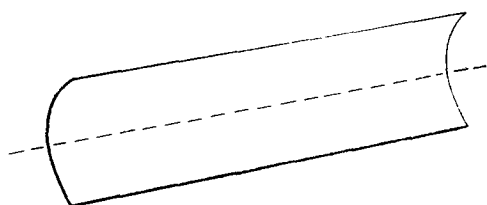
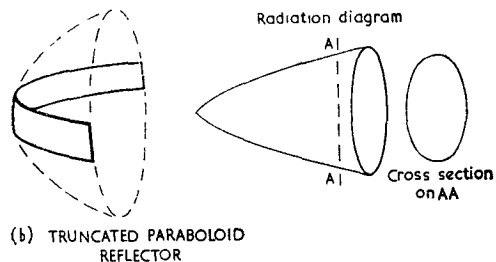


FIG. 9 PARABOLIC CYLINDER



(b) TRUNCATED PARABOLOID REFLECTOR

FIG. 8 TYPES OF PARABOLIC REFLECTOR

The beam shape may be modified by cutting away parts of the paraboloid and making an "orange peel" or "truncated paraboloid" as shown in Fig. 8a and b respectively. Note that as the reflector is narrow in the vertical plane and wider in the horizontal plane the beam formed is wide in the vertical plane and narrower in the horizontal plane.

The parabolic cylinder shown in Fig. 9 is normally energised by a number of aerials placed along the broken line. The parabolic curvature causes narrowing of the beam in the vertical plane.

Fig. 10 shows a parabolic "cheese" reflector. This is a narrow parabolic cylinder enclosed by flat plates on either side. It produces a fan-shaped beam, narrow in the plane in which the cheese is wide and wide in the other plane.

The parabolic torus reflector shown in Fig. 11 is formed by moving the parabola AB along the arc of a circle BC. The pencil-shaped beam formed by the reflector has fairly small side lobes. The torus reflector is normally used on ground radars where the reflector is stationary and scanning is achieved by moving the aerial feed.

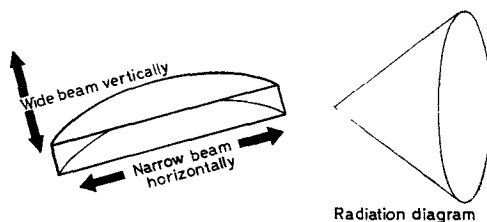


FIG. 10 PARABOLIC CHEESE REFLECTOR

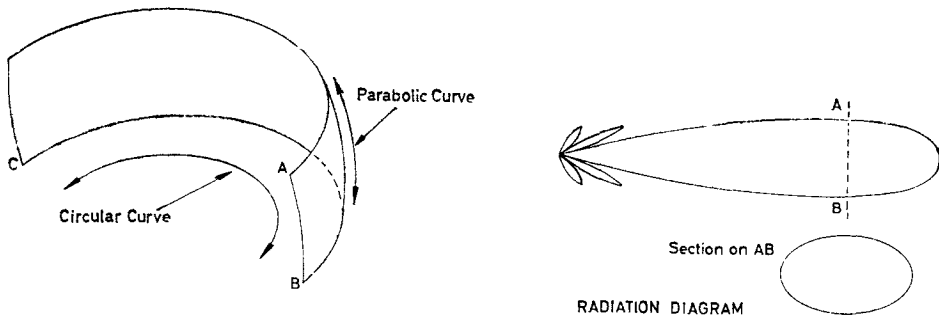


FIG. 11 PARABOLIC TORUS REFLECTOR

We have discussed some of the commonly used forms of parabolic reflector. They may be constructed of metal or metalized plastic and for airborne use are usually solid. Large ground installation reflectors are normally of metal mesh to reduce weight and wind resistance.

Cosecant-squared Reflector

The shape of the radiation pattern required from a radar aerial depends upon the type of job the radar has to perform. Special beam shapes can be produced by using specially-shaped reflectors.

An example is the *cosecant-squared* reflector used in some airborne search radars. A beam that is narrow in the horizontal plane (azimuth) but fairly wide in elevation is rotated horizontally so that the ground beneath the aircraft is illuminated by the radar energy. It is desirable for more energy to be beamed towards the ground that is further from the radar than towards that nearer the radar. Thus equal strength returns from similar ground targets at different ranges are obtained and the p.p.i. display is evenly illuminated.

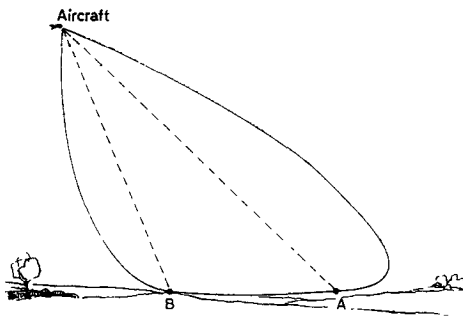


FIG. 12 COSECANT-SQUARED PATTERN

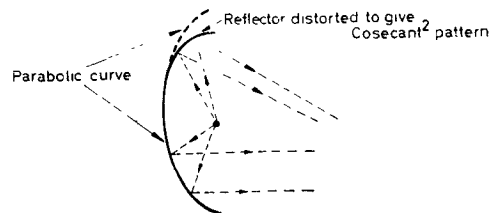


FIG. 13 COSECANT-SQUARED REFLECTOR

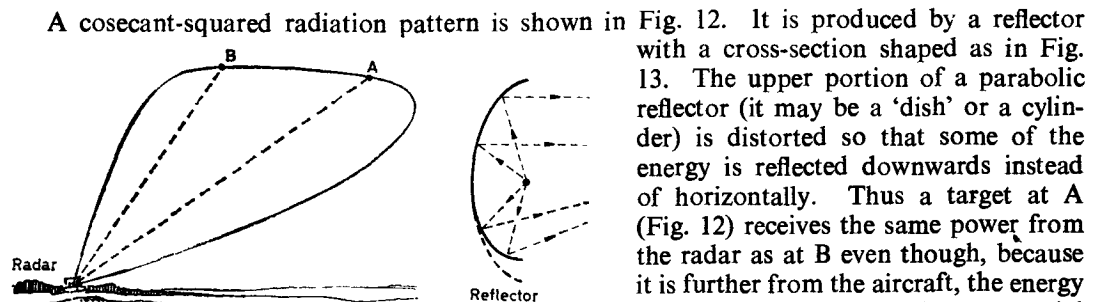


FIG. 14 EARLY WARNING COSECANT-SQUARED PATTERN

A cosecant-squared radiation pattern is shown in Fig. 12. It is produced by a reflector with a cross-section shaped as in Fig. 13. The upper portion of a parabolic reflector (it may be a 'dish' or a cylinder) is distorted so that some of the energy is reflected downwards instead of horizontally. Thus a target at A (Fig. 12) receives the same power from the radar as at B even though, because it is further from the aircraft, the energy is attenuated more. As the same aerial is used for receiving, constant illu-

(A.L. 2, August, 1965)

mination of the p.p.i. display results.

The same type of reflector, only this time inverted, is used in early warning radar so that similar targets at the same height but different angles of elevation return similar strength echoes (Fig. 14).

Methods of Feeding Reflector

Reflectors are usually energised by an aerial placed at the focal point, or in the case of a cylindrical reflector, from a number of aerials placed along the axis. It is arranged for most of the energy from the aerial to be directed into the reflector so that an almost parallel beam of energy is present at the aperture.

Dipole feed. At centimetric wavelengths the transmitter power is often fed down a waveguide inserted through the rear of the reflector (Fig. 15). A metal tongue fitted into the waveguide supports a dipole aerial and its reflector. The tongue splits the E field into two portions which excite each half of the dipole.

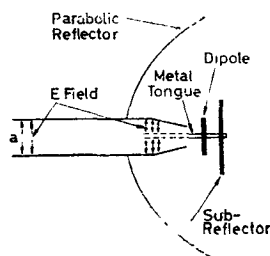


FIG. 15 WAVEGUIDE DIPOLE FEED

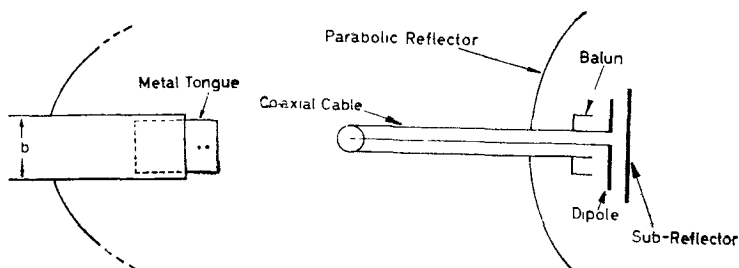


FIG. 16 COAXIAL DIPOLE FEED

At 10 cm and longer wavelengths the dipole may be fed from a coaxial line and a balanced-to-unbalanced device (a balun) is necessary (Fig. 16).

Rear feed with slots. In this case the waveguide is terminated in a resonant cavity in which two slots are cut to act as the radiating elements. If the slots are vertical the radiation is horizontally polarised (Fig. 17).

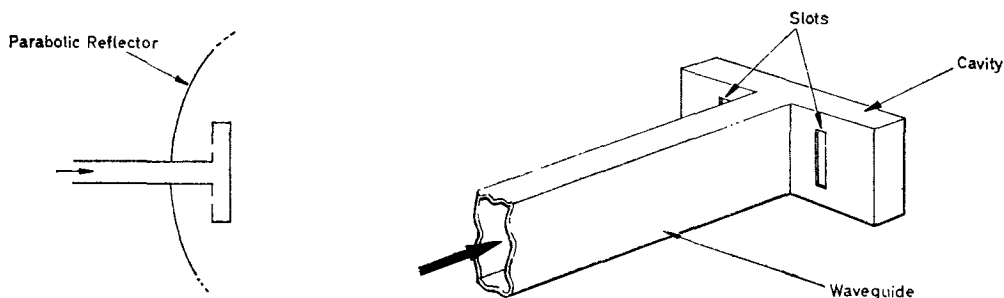


FIG. 17 SLOT FEED

Horn feed. If a waveguide is terminated by an open end (Fig. 18a) energy is radiated from the waveguide into space. However, the characteristic impedance of the waveguide will differ from that of free space and most of the energy will be reflected back down the guide, setting up standing waves. With a rectangular guide, propagating in the H_{01} mode, the characteristic impedance

$$(Z_H = 377 \frac{\lambda_g}{\lambda_a} \text{ ohms}) \text{ can be matched}$$

to the impedance of free space (377 ohms) by gradually increasing the b dimension of the guide (see p. 286).

By "flaring" the end of the waveguide to form a *horn* as in Fig. 18b maximum energy is radiated from the guide. As well as providing a good match, the horn is a directional radiator giving a fan-shaped beam. The guide may be flared in the H plane as shown in Fig. 18b when it is known as a *sectoral* horn or in both planes as in the *pyramidal* horn of Fig. 18c. By placing

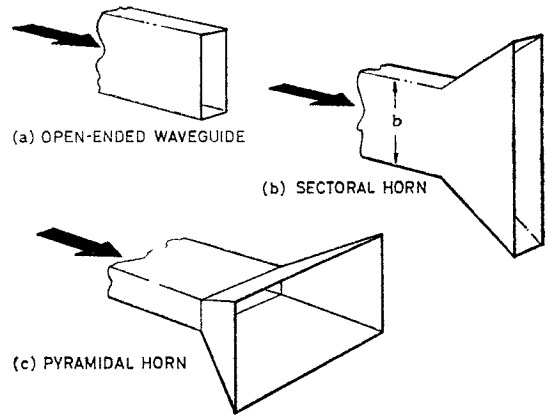


FIG. 18 WAVEGUIDE HORN RADIATORS

is known as a *sectoral* horn or in both planes as in the *pyramidal* horn of Fig. 18c. By placing

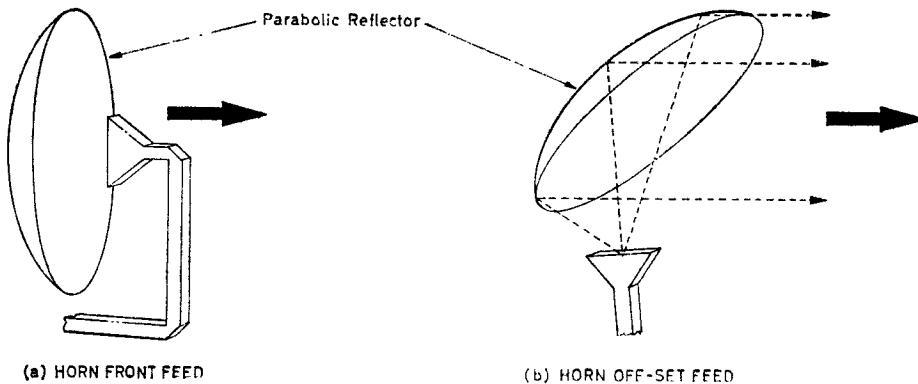


FIG. 19 TYPES OF HORN FEED

the horn at the focal point of a parabolic reflector (Fig. 19a) a very narrow beam is radiated.

The horn and its waveguide feed form an obstruction to the radiated beam. This distorts the radiation pattern and also affects the impedance of the horn. By using an off-set feed as shown in Fig. 19b these effects are avoided.

A development of the horn feed is shown in Fig. 20 and is known as the *hohorn*. Here the end of the waveguide is shaped so that the side AB forms a parabola. Energy from the waveguide is reflected from the parabola through the aperture BC. The hohorn feed provides a fan-shaped beam narrow in one plane and wide in the other. It is often used to feed a cheese reflector and when used as shown in Fig. 21 the aerial produces a horizontally polarised beam narrow in azimuth and wide in elevation.

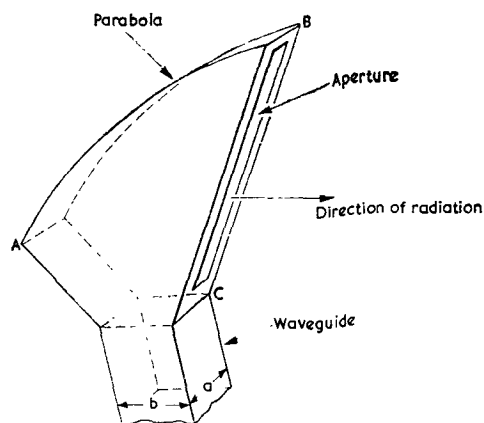


FIG. 20 THE HOGHORN

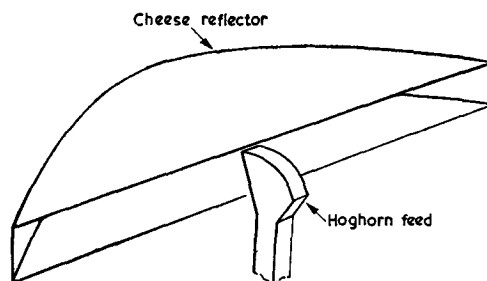


FIG. 21 HOGHORN FEED TO A CHEESE REFLECTOR

The Cassegrain Aerial

The aerial shown in Fig. 22 is developed from a principle used to reduce the length of optical telescopes and is known as the *Cassegrain* aerial. A parabolic reflector is fed from the rear

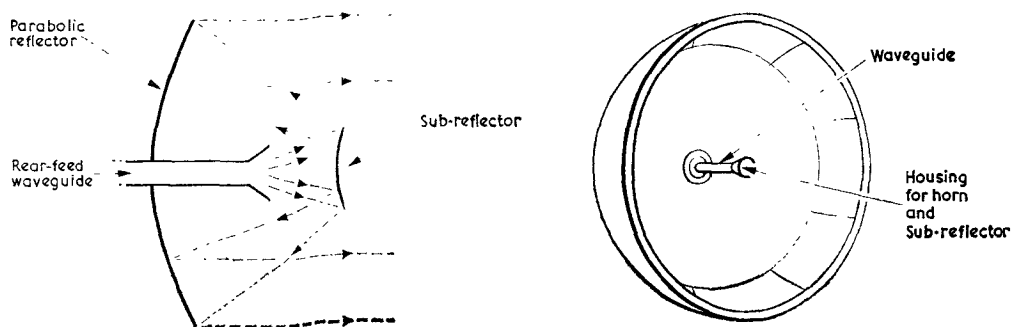


FIG. 22 CASSEGRAIN AERIAL

by a waveguide horn. The waveguide energy is reflected from a small convex sub-reflector into the main parabolic reflector. This 'double-focusing' results in a very narrow beam from an aerial of much smaller dimensions than those required for a normal horn and reflector.

As well as the advantages of smaller size, lower cost and simpler mounting, the rear feed means that a shorter length of waveguide between aerial and radar is required. Thus losses and noise due to waveguide wall resistance are less and a low-noise amplifier such as a parametric amplifier or maser (see Chapters 9 and 10) can be easily mounted directly behind the main reflector.

Multi-beam Reflector Aerials

When a single reflector is fed with energy from several horn aerials mounted one above the other, a number of beams at different angles of elevation are formed. This type of aerial can be

used to produce beams giving both low- and high-angle coverage for a surveillance radar. By forming several beams, each beam requiring a separate horn, the aerial can be used as a height finder. The outline of a 'stacked horn' aerial in which four horns feed into a parabolic reflector is shown in Fig. 23. The radiation diagram shows that four fixed beams at different angles of elevation are formed.

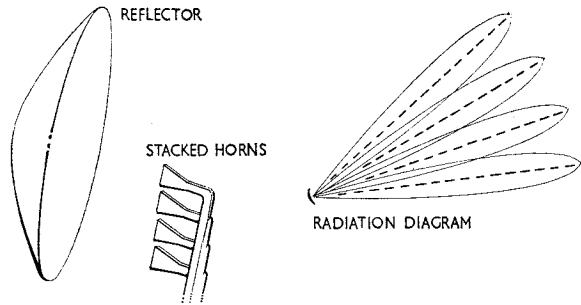


FIG 23. STACKED HORN AERIAL

Large ground radars are sometimes *scanned* by moving the position of the feed point and keeping the reflector stationary. This avoids the need to move a large metal reflector and simplifies the scanning mechanism. By moving the feed point away from the focal point of a parabolic reflector the direction in which the beam points is changed (Fig. 24) and the aerial scans over a limited sector. If the scanning angle is made too great the beam becomes distorted and the aerial gain falls.

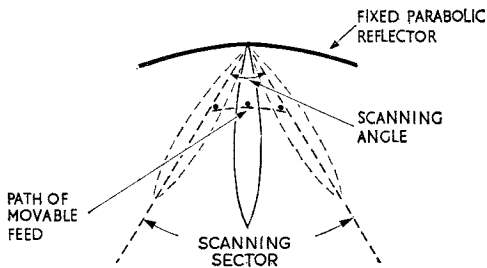


FIG 24. MOVABLE FEED POINT SCANNING

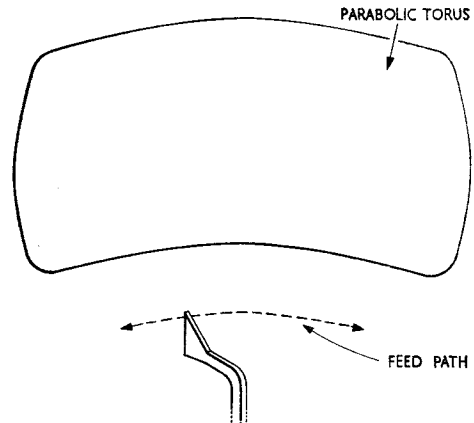


FIG 25. PARABOLIC-TORUS REFLECTOR WITH VARIABLE FEED POINT SCANNING

The parabolic-torus reflector (page 320) can be used to provide much wider scanning angles (up to 120°) than the simple parabolic reflector, without distorting the beam. A single horn may be moved over the arc of a circle as shown in Fig. 25 or the transmitter output may be switched to a number of fixed feed points by means of the "organ pipe" scanner shown in Fig. 26. In this type of scanner a single input horn is rotated so that it feeds energy to each output horn in succession. Each waveguide length in the organ pipe scanner is equal and so the feed horn is effectively moved across the reflector causing the beam to scan horizontally.

The Quarter-wave Plate

At centimetric wavelengths, rain droplets cause echoes which show on the radar display as *clutter*, obscuring the real targets. One method of overcoming this effect is to radiate circularly polarised waves. The radar is then sensitive to the difference between echoes from real targets and those caused by rain, and the clutter can be removed. We have already seen that the turn-

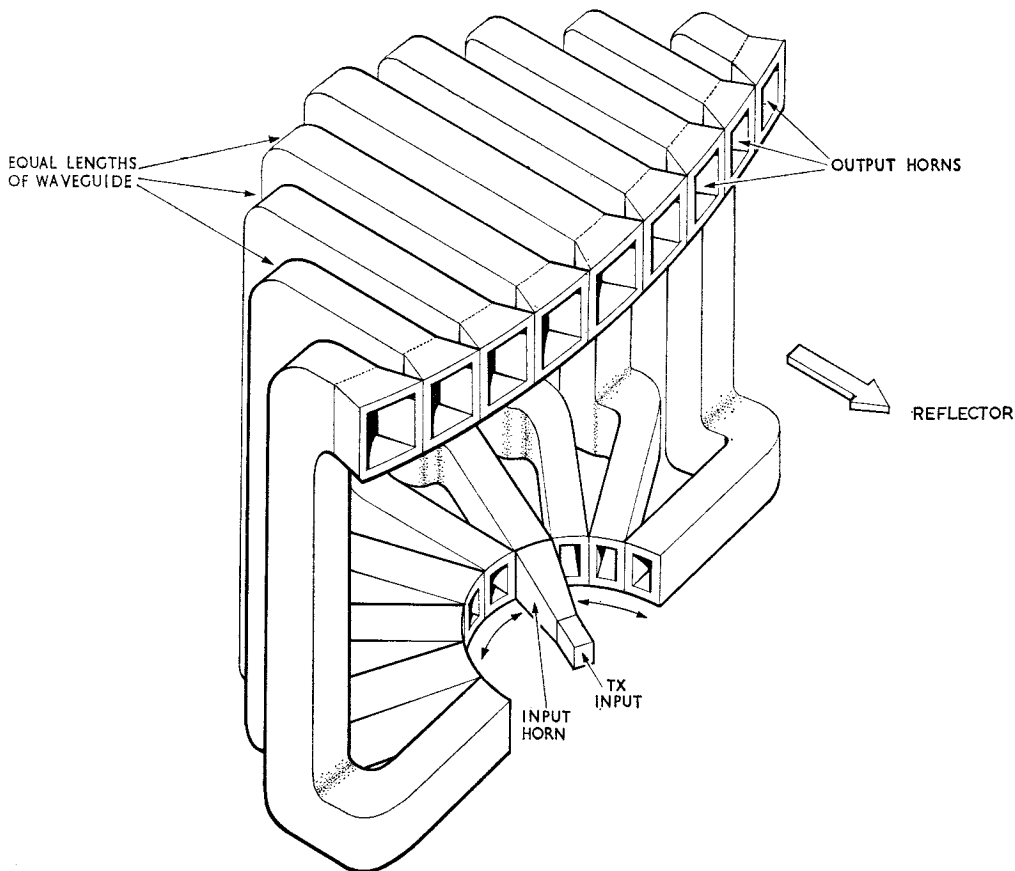


FIG 26. ORGAN-PIPE SCANNER

stile junction (p. 301) may be used to change linear polarisation into circular polarisation. Another method is to place an assembly of parallel plates, known as a *quarter-wave plate*, between the aerial and reflector.

As shown in Fig. 27a the plates are mounted so that the E field vector is at 45° to the plates. The horizontal component of the E field is unaffected as it passes through the plates but the vertical component propagates as if it were in a waveguide. The phase velocity of a wave in a

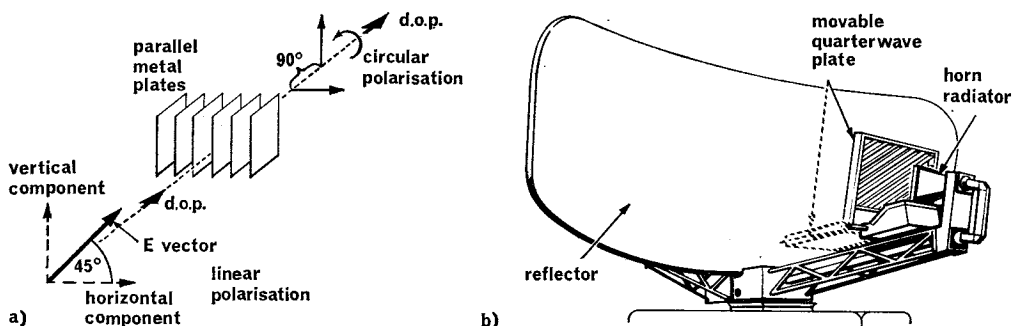


FIG 27. THE QUARTER-WAVE PLATE

waveguide is greater than the velocity in free space so the vertical component is speeded up relative to the horizontal component. The plate spacing and size are such that the vertical component is advanced in phase by 90° , or a quarter of a wavelength, hence the name of the device. A circularly polarised wave can be considered as consisting of two linearly polarised waves travelling in the same direction but with their E fields at right angles to each other and 90° out of phase. Thus the wave radiated from the reflector is circularly polarised.

Fig. 27b shows a quarter wave (or anti-rain) plate mounted between a horn radiator and reflector. The quarter-wave plate assembly can be lowered in fine weather; the polarisation is then horizontal instead of circular.

Lens Aerials

We have seen that a parabolic "mirror" can be used at centimetric wavelengths to form a narrow beam, i.e. it can focus centimetric waves. A *lens* is an optical device used to focus light rays and, as with the parabolic mirror, the lens can be used at centimetric wavelengths to form a narrow beam of e.m. energy.

Dielectric lens. A centimetric lens may be made from a dielectric material such as polystyrene or polyethylene. These materials have the property of *slowing down* rays which pass through them. The lens is shaped such that diverging rays are bent in the lens and emerge as an almost parallel beam (Fig. 28). To achieve this the centre of the lens (AB) must be thicker than the edges (C and D).

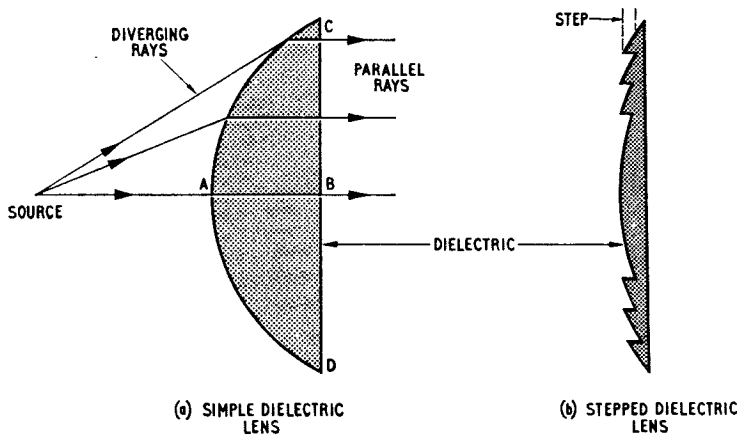
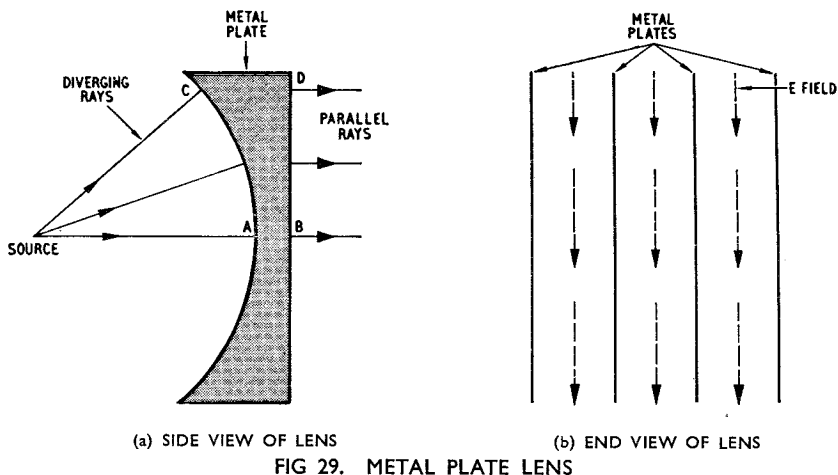


FIG 28. DIELECTRIC LENSES

The simple dielectric lens of Fig. 28a may be quite thick and hence heavy and bulky. It can be made lighter by cutting steps in the curved surface as shown in Fig. 28b. If the thickness of the step is made a whole number of wavelengths long the action of the lens is unaffected.

Metal plate lens. Another form of microwave lens is constructed of a number of shaped metal plates. In the *metal plate lens* the ray is *speeded up* as it passes through the lens. Thus to produce a parallel beam from diverging rays the plates forming the lens must be thicker at the edges CD than in the centre AB (Fig. 29a).

The metal plate lens consists of several conducting strips placed parallel to the E field and spaced slightly more than half a wavelength apart (Fig. 29b). To the wave the lens acts as a number of waveguides in parallel. We know that the phase velocity in a waveguide is greater than the velocity in free space (p. 285); by shaping the plates as in Fig. 29a the diverging rays from the



source are speeded up more than the rays at the axis of the lens. Thus the rays emerge from the lens to form a parallel in-phase beam.

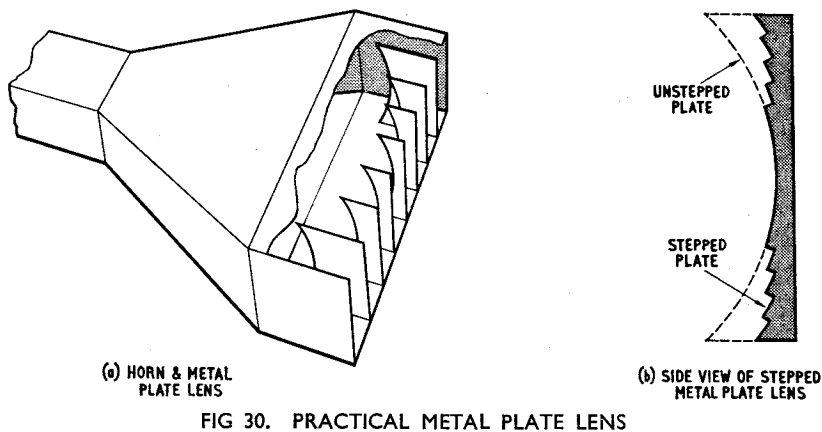
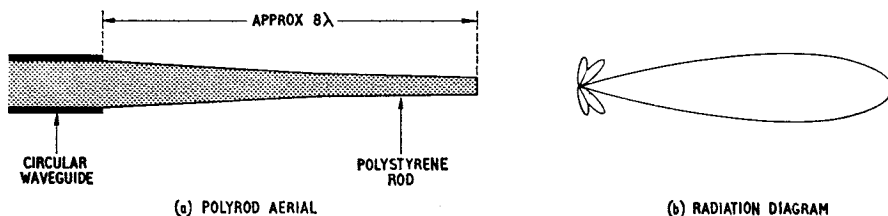


Fig. 30a shows a metal plate lens used to improve the radiation pattern of a waveguide horn aerial. The lens and horn form a very narrow pencil-shaped beam with small side-lobes. To reduce weight and size, the plates of the lens are usually stepped as shown in Fig. 30b.

Polyrod Aerial

A length of tapered polystyrene inserted into the end of a circular waveguide (Fig. 31a) acts as an end-fire directional aerial. The energy tends to follow the dielectric but, due to the



taper, is transferred into space, forming a pencil-shaped beam as in Fig. 31b.

Slotted Waveguide Arrays

We have seen in Section 3 that when a number of aerials are spaced half a wavelength apart and fed in phase, energy radiated from the aerials adds in some directions and cancels in others, depending upon the phase relationships. This type of aerial array is the linear *broadside* array and gives a beam at right angles to the line of the aerials (Fig. 32).

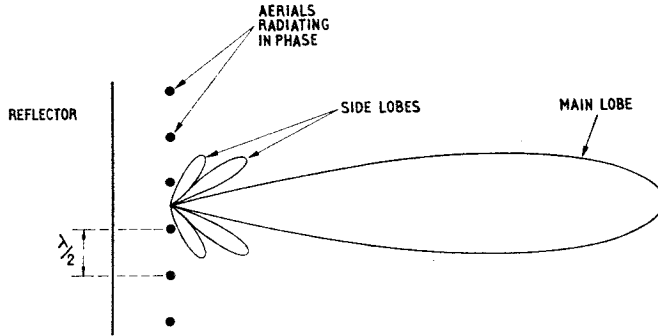


FIG 32. THE BROADSIDE ARRAY

In order to obtain this type of radiation pattern each aerial must radiate an *equal* amount of *in-phase* energy and the aerial spacing must be half a wavelength. We have seen that if the phase of the energy radiated by adjacent elements in the array is altered the radiated beam will point in a direction other than at 90° to the line of the aerials.

On page 290 we saw that slots cut at certain positions in a waveguide wall will radiate energy from the guide. If these slots are made half a wavelength long they behave as half-wave dipoles. Thus by cutting a series of half-wave slots in the wall of a waveguide and arranging that they each radiate *equal* amounts of *in-phase* energy we have a microwave broadside array.

Displaced Slots

A slot cut along the centre line of the broad side of a waveguide (slot 1 in Fig. 33a) will not radiate energy since it does not interrupt the flow of wall current. However, if the slot is displaced slightly (slot 2) it interrupts a small wall current and radiates. As the displacement is

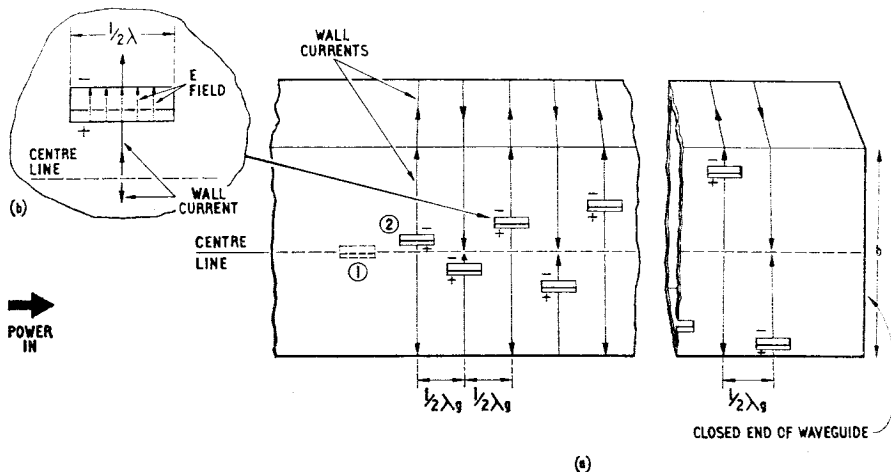


FIG 33. DISPLACED SLOTS IN A WAVEGUIDE

increased, coupling increases and a greater percentage of the energy in the guide is radiated. With a long waveguide array the displacement is gradually increased so that as the energy in the guide decreases due to slot radiation, the coupling increases to give equal radiation from each slot. Thus we have satisfied one requirement of a broadside array.

The wall current reverses phase every half a guide wavelength and if we space the slots $\frac{1}{2}\lambda_g$ apart on the *same* side of the centre line adjacent slots would radiate *in anti-phase* giving zero radiation at right angles to the guide. Therefore we *alternate* the slots about the centre line (Fig. 33a); all the slots now radiate in-phase energy and so give a narrow beam at 90° to the waveguide.

As shown in Fig. 33b the E field is formed parallel to the narrow sides of the slots and therefore when the waveguide is mounted horizontally this slotted waveguide array gives vertical polarisation. The beam produced is broad in elevation and narrow in azimuth. If the number of slots is increased, i.e. if the array is made longer, the azimuth beam width is decreased.

Inclined Slots

When a half wavelength slot is cut in the narrow a dimension of a waveguide such that it interrupts the flow of wall current, it will act as an aerial.

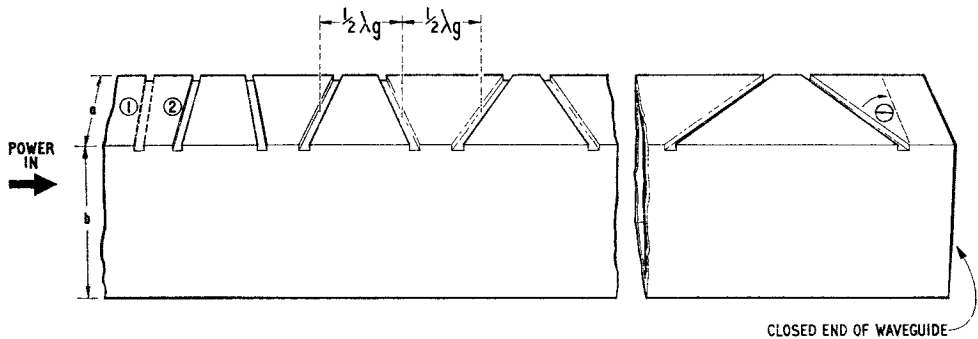


FIG 34. INCLINED WAVEGUIDE SLOTS

Slot 1 in Fig. 34 lies parallel to the wall current and therefore does not radiate. Slot 2 is inclined slightly and it will radiate. As the angle of inclination (θ) is increased the coupling increases. With an array of slots this angle is progressively increased to compensate for the reduction of energy in the guide so that each slot radiates equal energy.

If the slots are spaced $\frac{1}{2}\lambda_g$ apart and inclined *in the same* direction adjacent slots radiate in anti-phase and radiation at right angles to the array is zero. By *alternating* the slopes of adjacent slots as in Fig. 34 the slots radiate in phase and a beam at 90° to the line of the array is formed.

The slots are made half a wavelength long and therefore they are continued into the broad sides of the guide. A longer array with more slots results in a narrower beam in the plane of the array. For example, an array with 79 slots gives a beam width of $1\frac{1}{2}^\circ$. The inclined slot array gives polarisation in the same direction as the length of waveguide.

Resonant Slotted Array

So far we have considered arrays in which the slot spacing is $0.5\lambda_g$. With this spacing a beam is formed at 90° to the line of the array. This can be seen from Fig. 35a which shows two slots of a slotted waveguide array. The time taken for the field pattern in the guide to move from slot A to slot B ($0.5\lambda_g$) at the phase velocity is the same as for the radiated wave to travel from A to P ($0.5\lambda_g$) at free space velocity. As adjacent slots are reverse coupled, radiation is therefore in phase along PP_1 and the beam points at 90° to the line of slots (Fig. 35b).

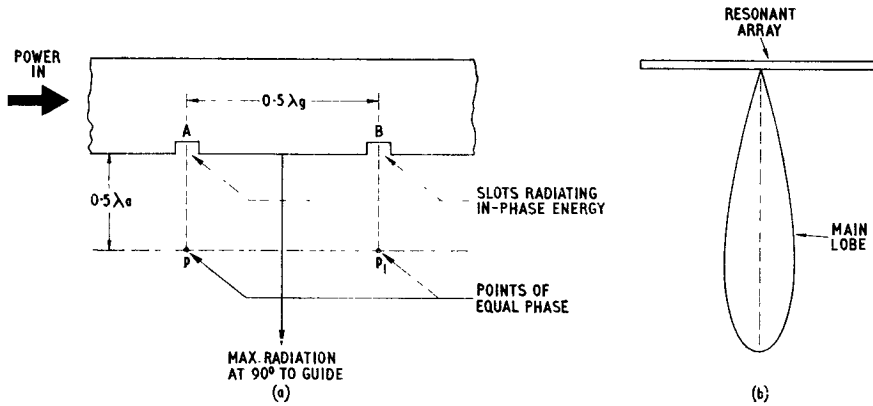


FIG. 35 RESONANT ARRAY

This arrangement is called a *resonant array*. As each slot radiates only a fraction of the total energy in the guide, individual slots are not matched to the guide and each slot sets up a standing wave inside the guide. As the slots are spaced $0.5\lambda_g$ apart the standing waves are additive and give a large standing wave at the input to the array. This presents an impedance which must be matched to the main guide. With a long array a slight change in transmitter frequency causes a large change of impedance and the matching is no longer correct. In other words the resonant array has a very *narrow bandwidth*.

Non-resonant Slotted Array

By spacing the slots slightly greater than $0.5\lambda_g$ apart (say $0.55\lambda_g$) the standing waves due to each slot are no longer additive but tend to cancel in a long array. Matching is less critical and the bandwidth wider. This is the advantage of the *non-resonant array*.

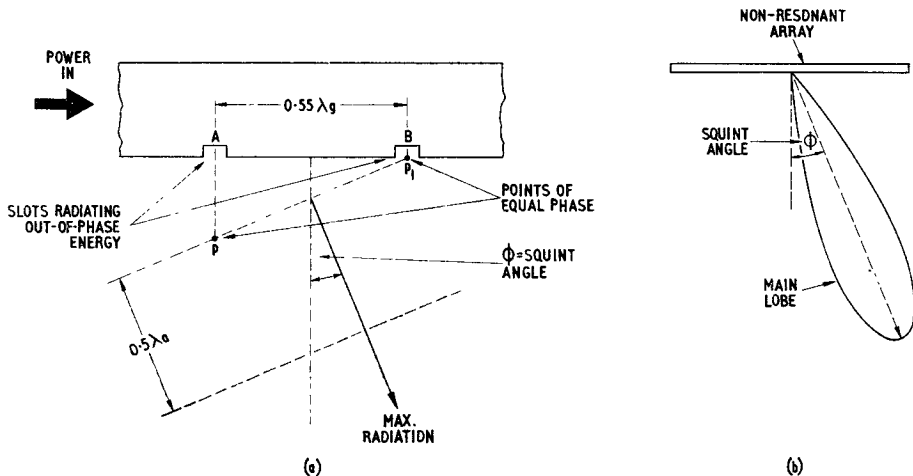


FIG. 36 NON-RESONANT ARRAY

When the slots are spaced $0.55\lambda_g$ apart (Fig. 36a) the time taken for the field pattern in the guide to move from slot A to slot B is greater than the half-cycle time. Therefore the energy radiated from A will have reached P by the time the same wavefront is radiated at the reverse-coupled slot B. The two radiations will therefore be in phase along line PP_1 . Maximum radiation at right angles to this line will be at a squint angle ϕ° (Fig. 36b).

The squint angle depends upon the guide wavelength, i.e. upon the frequency and waveguide dimensions, and also upon the slot spacing. The squint angle is not necessarily a disadvantage. It can easily be compensated for and in some radars it is used to tilt the beam in the required direction. We shall see later that by continuously varying the phase of the energy fed to each slot the beam angle can be continuously changed, i.e. the slotted waveguide array can scan without either the array or its reflector being moved.

Practical Slotted Waveguide Array

The slotted waveguide array is normally used with a parabolic or cosecant-squared reflector to give the required beam shape. Fig. 37 shows a typical arrangement for an S band (10 cm) ground radar. The horizontal non-resonant waveguide array has inclined slots cut in the narrow

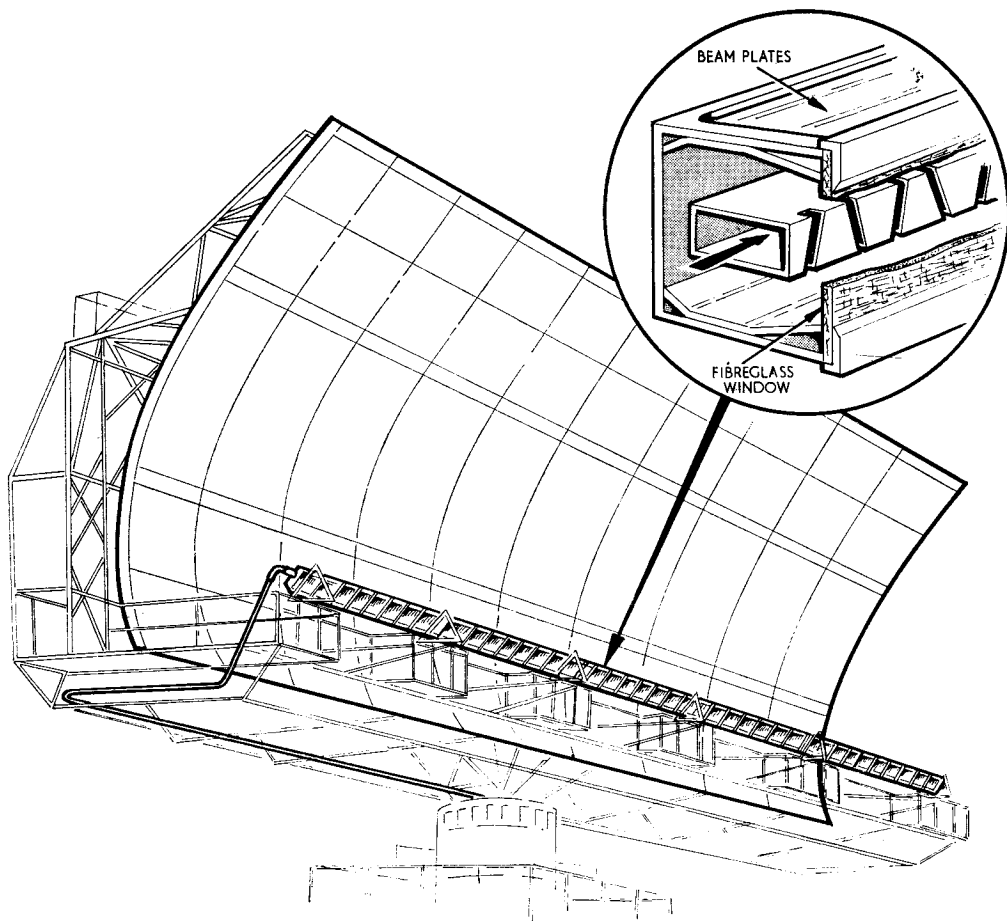


FIG. 37 TYPICAL GROUND RADAR SLOTTED WAVEGUIDE ARRAY WITH REFLECTOR

dimension facing into a cosecant-squared cylinder. The waveguide radiation produces a beam about 0.3° wide in the horizontal plane while the reflector forms a cosecant-squared pattern in the vertical plane.

As shown in the inset of Fig. 37 the waveguide array is enclosed in a channel, the top and bottom plates of which beam the radiation from the array so that the reflector is correctly illuminated. The front of the channel is fitted with a fibreglass window. This window is transparent

to e.m. waves but makes the channel airtight so that dry air under pressure may be pumped through the waveguide to prevent corrosion, arcing or corona discharge.

The slots radiate approximately 90 per cent of the total energy in the guide, the remainder being absorbed by a resistive dummy load at the end of the array remote from the transmitter input. The squint angle is about 4° and is compensated for by mounting the waveguide array such that the feed end is further from the reflector than the absorber end. Horizontal polarisation is used but if vertical polarisation is required, displaced slots cut in the broad dimension could be used.

Beam Steering

We have seen that if the distance between slots in a waveguide array is altered the phase of the energy radiated from adjacent slots is changed and the beam formed by the array points in a different direction. By arranging that the phase difference between adjacent slots varies *continuously* the beam can be continuously steered (or scanned) without moving the aerial array or its reflector.

One method of obtaining a relative phase change between adjacent slots is by altering the guide wavelength of the energy in the guide. In Chapter 5 (p. 284) it was explained that by *reducing* the broad dimension of the guide, λ_g is *increased*; if the broad dimension is *increased*, λ_g is *decreased*. Thus by providing a slotted waveguide array with a movable wall which can be pushed in and out by a motorised drive the beam can be continuously steered from side to side.

The principle is illustrated in Fig. 38. With the broad dimension at its normal size ($b = 0.707 \lambda_a$) the slots, spaced $0.5 \lambda_g$ apart radiate in-phase energy and the beam points at right angles to the line of slots (Fig. 38a). As the b dimension is decreased λ_g increases and the slots, now spaced *less* than $0.5 \lambda_g$ apart, radiate out of phase. The beam swings to one side of the normal (Fig. 38b). By increasing the b dimension above its normal size λ_g decreases and the slot spacing becomes *greater* than $0.5 \lambda_g$. The beam is steered in the opposite direction (Fig. 38c).

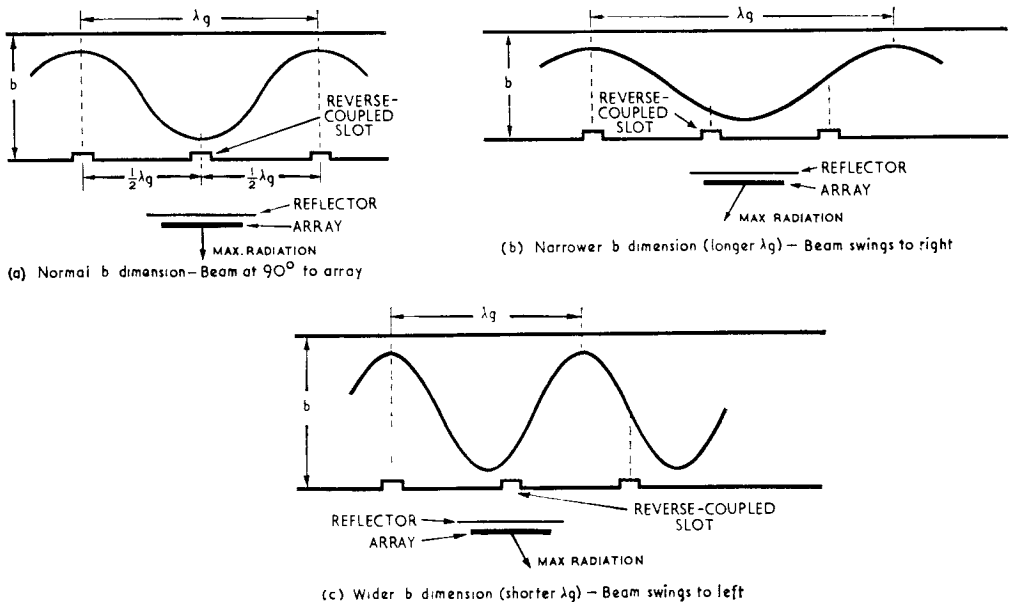


FIG. 38 BEAM STEERING BY CHANGING GUIDE DIMENSION

Another method of steering the beam of a slotted waveguide array is by mounting a movable iris near each slot. A change in position of the irises changes the phase of the energy radiated from the slots, thus swinging the beam from side to side.

By varying the d.c. magnetising field of ferrite phase shifters mounted inside the waveguide array the phase of energy radiated by a slot or group of slots can be varied relative to other slots or groups of slots in the array. There is no mechanical movement of any kind and the beam is *electronically steered*.

Another method of electronically steering a waveguide array makes use of the fact that when a waveguide is terminated with a pure reactance the phase of the reflected energy depends upon the value of terminating reactance. Thus by varying the value of a capacitance mounted in a waveguide directional coupler fitted between the radar transmitter and waveguide array the phase of energy fed to the array is varied and the beam can be steered.

A *varactor* is a semiconductor diode whose capacitance can be varied by varying the value of applied bias (see p. 344). Fig. 39 shows a suitable arrangement using a short-slot directional

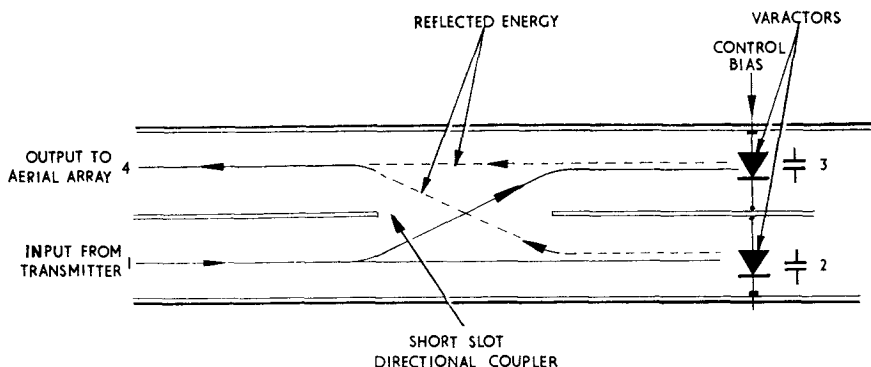


FIG. 39 USE OF VARACTORS IN ELECTRONIC BEAM STEERING

coupler and two varactors. The transmitter output is fed into arm 1 and divides equally into arms 2 and 3. The two signals are reflected by the varactors with a change of phase depending upon the values of capacitance. The reflected signals add together in the output arm and are fed to the aerial array. By varying the bias on the varactors the capacitance, and hence the phase shift, is varied and the beam can be steered.

This method of beam steering is fast and requires only one control signal. The varactors are compact and they absorb very little power.

A waveguide array may be electronically steered by changing the *frequency* of energy fed to the array. A change of frequency alters the guide wavelength (λ_g) and thus causes a change in phase difference between adjacent slots in the array. The beam then points in a different direction. By sweeping the transmitter frequency over a given band the beam is continuously steered over a certain angle of scan.

The electronic and mechanical methods of beam steering outlined enable the beam to be scanned without moving the aerial structure. Scanning speeds are greatly increased and power requirements for normal turning gear are greatly reduced or eliminated. The main disadvantage is that the scanning angle is limited to approximately $\pm 30^\circ$.

Applications of Centimetric Aerials

We have seen how various types of centimetric aerial can produce beams of different shapes. We shall now discuss some of the uses of these aerials.

Search in azimuth. A radar which is required to provide the bearing of a target needs an aerial which gives a beam narrow in the horizontal plane but wider in the vertical plane. At

centimetric wavelengths a slotted waveguide array feeding into a cosecant-squared reflector will provide such a beam. The array gives a beam narrow in azimuth and the reflector shapes the vertical pattern so that targets at various angles of elevation can be detected.

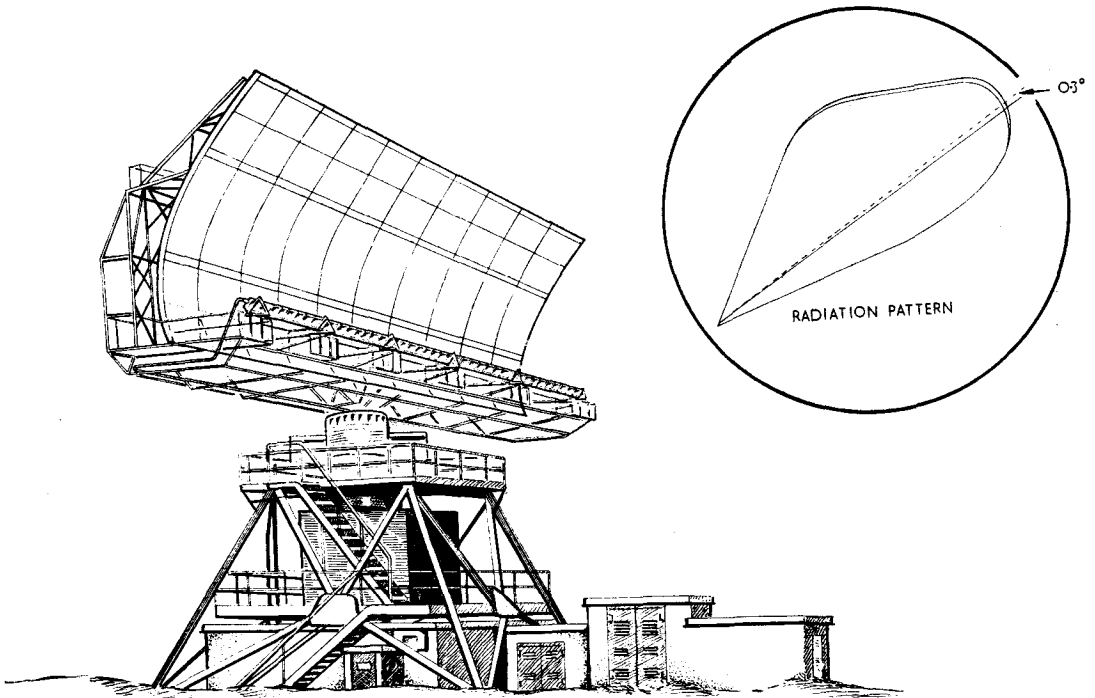


FIG. 40 GROUND SURVEILLANCE RADAR AERIAL

Fig. 40 shows the aerial of a ground surveillance radar working in the S (10 cm) band. The array and its reflector are mounted on a motor-driven turntable which rotates the aerial assembly through the required search sector. An airborne search radar for use as an aid to navigation and blind bombing and working in the X (3 cm) band would employ a similar aerial system but because of the shorter wavelength the assembly would be much smaller and lighter.

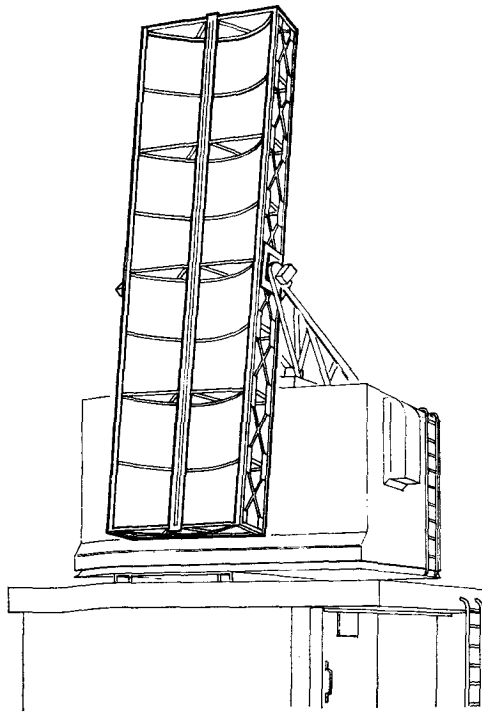
Height-finding aerials. For finding the height of a target, its range and elevation angle must be determined. A number of different types of aerial can be used for this purpose. Fig. 41 shows two aerials which produce a "beaver tail" beam narrow in elevation and wider in azimuth.

This beam is usually directed onto the azimuth bearing of the target by information obtained from the surveillance radar. The aerial is then made to nod up and down so that the beam passes through the target. From the range and elevation angle so obtained the target height can be computed.

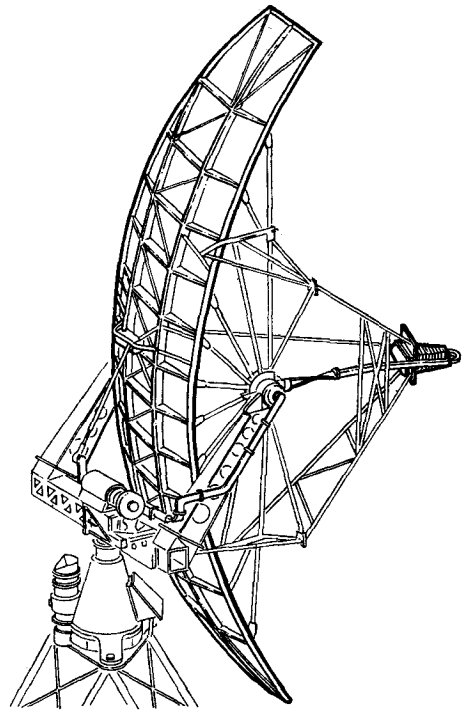
Fig. 41a shows a vertical slotted waveguide array with a parabolic cylinder reflector and Fig. 41b shows a horn-fed truncated parabolic aerial; both are used as height finders.

Another method of height-finding employs an aerial which produces a stack of narrow pencil-shaped beams at various fixed angles of elevation. The target response from each beam is fed into a separate receiver and when the receiver outputs are compared an accurate indication of the target's elevation angle is obtained (see p. 246).

An aerial suitable for this type of height finding is the stacked horn aerial feeding into a parabolic reflector (see Fig. 23). As the aerial is turned in azimuth the whole pattern rotates until the target is detected.

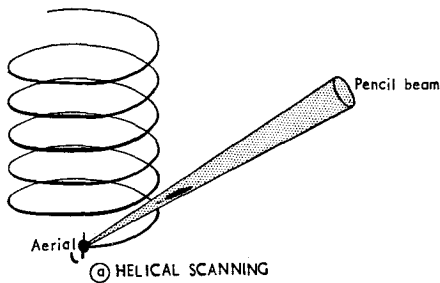


(a) SLOTTED WAVEGUIDE WITH PARABOLIC REFLECTOR

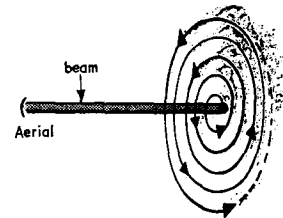


(b) HORN-FED TRUNCATED PARABOLIC

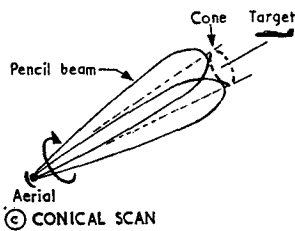
FIG. 41 HEIGHT-FINDING AERIALS



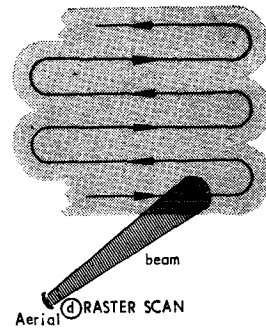
(a) HELICAL SCANNING



(b) SPIRAL SCANNING



(c) CONICAL SCAN



(d) RASTER SCAN

FIG. 42 TYPES OF PENCIL BEAM SCANNING PATTERNS

Pencil-shaped beams. A parabolic or "dish" reflector may be used to form a pencil-shaped beam. Different types of parabolic reflector (for example, the orange peel and truncated paraboloid reflectors) can be used to alter the shape of the beam. A beam narrow in both planes must be steered or scanned in order to search a required volume of space. The beam must follow a regular pattern in order that all the required space is searched. There are many types of scanning pattern, each suited to a particular task. Four common patterns are shown in Fig. 42. As explained in Section 1 (p. 36) the scanning speed must be related to the radar p.r.f. to ensure that targets are illuminated by several radar pulses.

Pencil-shaped beams can be moved by mounting the aerial and reflector on a platform which can be turned and tilted over the required angles of azimuth and elevation. In airborne radars movement of the platform due to aircraft pitch and roll can be compensated for by gyroscopic stabilisation.

Scanning the beam may also be achieved by moving the aerial feed point and by mechanical or electronic beam steering methods already considered.

Monopulse radar. Conical scanning (Fig. 42c) requires information from at least four pulses to derive an error signal with which to position the aerial in azimuth and elevation. If the amplitude of these pulses varies due to varying attitudes of the target or to intentional jamming (see Section 6), the accuracy of azimuth and elevation information is weakened. In a *monopulse* tracking radar the angular measurements are made using *one pulse only*, and therefore these inaccuracies cannot occur. Two or more beams are formed *simultaneously* and by comparing the relative phase or amplitude of the echo pulse received in each beam the angular position of the target can be determined from information contained in a single pulse.

Fig. 43a shows an X band (3 cm) monopulse aerial used in an AI automatic tracking radar.

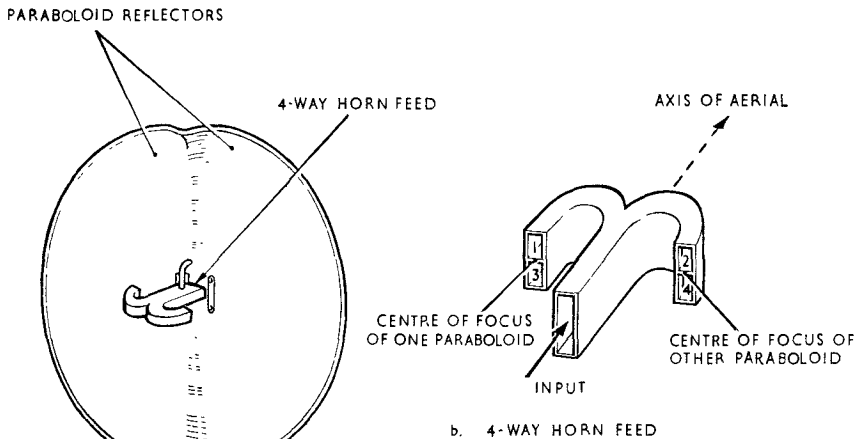


FIG. 43 MONOPULSE MULTI-BEAM AERIAL

The reflector, a double paraboloid, is fed from the rear by a four-way horn. This feed has four apertures, each pair of which is aligned with the focus of each paraboloid (Fig. 43b).

The radiation diagram in elevation shows two tilted beams which cross. Thus if a target lies on the axis of the aerial, the *amplitude* of signals received at apertures 1 and 2 will be the same as that of signals received at apertures 3 and 4. If the target lies off the vertical axis of the aerial there will be a difference in the amplitude of signals received by the two pairs of apertures. The difference signal is used to align the aerial, in elevation, to the target. Azimuth bearing of the target is found by comparing the difference in *phase* between signals received by apertures 1 and 3, and 2 and 4. This difference is converted into a signal which aligns the aerial in azimuth.

Radomes

Microwave aerals carried by aircraft and missiles are usually mounted inside a special dome, shaped so that aerodynamic properties are not affected. These *radomes* are made of a material such as fibreglass or hycar which does not reflect, refract or absorb the radiated energy. The radome may be built into the nose or tail of an aircraft (Fig. 44a and b) or may be a blister-shape in the fuselage (Fig. 44c). It is often removable so that the radar can be serviced from outside the aircraft. It is important that the radome is not painted with a metal-base paint.

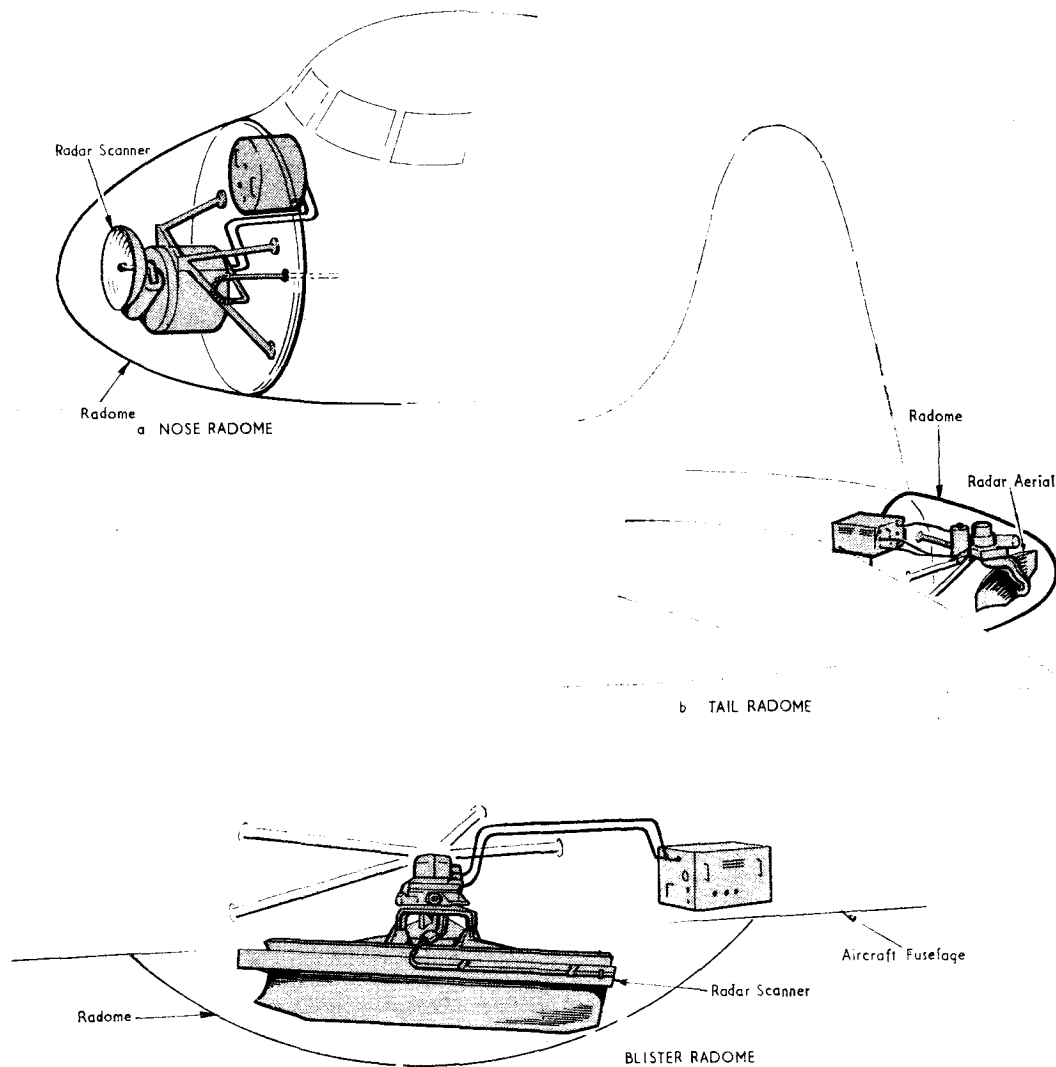


FIG. 44 TYPES OF AIRCRAFT RADOME

Ground radar aerals are sometimes enclosed in large radomes to protect them from the effects of ice and high winds. The radome is usually spherical and may be of rigid plastic for use at longer wavelengths (Fig. 45). Radomes designed for ground radars working above 10^6 c/s are usually made of a flexible airtight material which is kept rigid by air pressure from within.

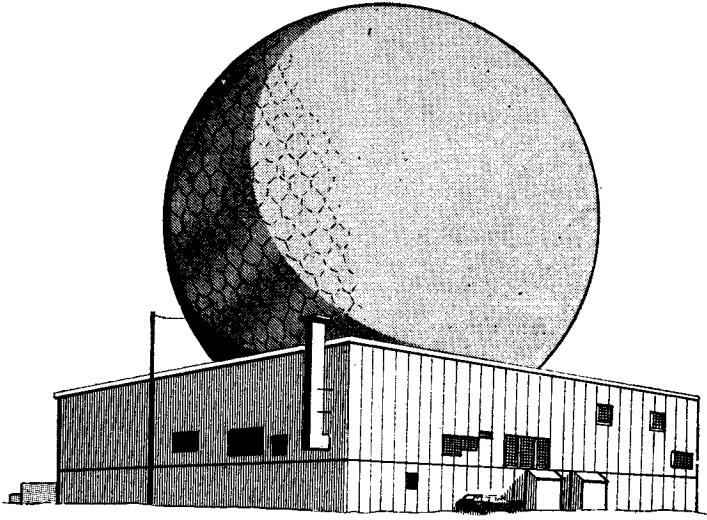


FIG. 45 RIGID GROUND RADAR RADOME

MICROWAVE SEMICONDUCTOR DEVICES

Introduction

In previous chapters of this section we have seen how thermionic devices such as the klystron, magnetron, t.w.t., etc., have been developed to generate and amplify microwave energy. These 'hot cathode' devices generate considerable noise and where they can be replaced by semiconductor devices the noise factor is considerably improved. In addition, semiconductors have the following advantages over thermionic devices:—

- a. They are smaller, lighter and more reliable.
- b. Higher efficiencies are possible—less energy is wasted as heat.
- c. Smaller, more compact power supplies can be used to operate them.

Because of its low noise and low transit time properties the point-contact semiconductor diode has for many years been used in the first stage of a microwave radar receiver. More recently, other microwave semiconductor devices have been developed and future radars will undoubtedly make use of these devices to improve performance.

This chapter deals with the properties and microwave applications of some of these devices and outlines the principles of low-noise parametric amplification and frequency conversion.

The Point-contact Diode

The point-contact diode consists of a metal "whisker" which presses on a small crystal of p- or n-type semiconductor (Fig. 1a). A diode made of p-type silicon with a tungsten whisker can handle very small signals but it is easily damaged by high power. In some cases welded

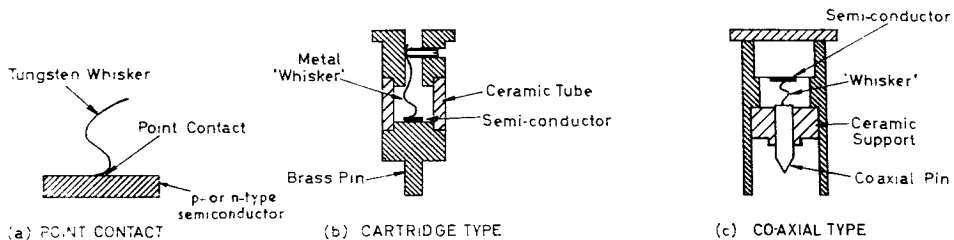


FIG 1. POINT-CONTACT DIODE

or gold bonded contacts are used. Crystals of *gallium arsenide* can withstand high temperatures and can operate at frequencies up to 100 Gc/s.

Two forms of crystal diode construction used in different types of waveguide mount are shown in Fig. 1b and c. The cartridge type is usually mounted inside the guide or coaxial cable and the coaxial type, which is commonly used above 3 Gc/s, is mounted outside the waveguide and fed via a short length of coaxial cable.

The typical diode characteristic shown in Fig. 2a illustrates the rectifying properties of a point-contact diode and Fig. 2b shows its equivalent circuit. The resistance r is the variable

“barrier” resistance of the contact and varies with the applied voltage thus enabling the diode to be used as a mixer. The contact capacitance C also varies with the applied voltage and is very low (approximately 0.2 pF) so that at microwave frequencies it does not short-circuit r . The series resistance R is the constant ohmic resistance of the semiconductor.

Point-contact diodes are fairly robust but when used as microwave mixers their life depends upon the effectiveness of the TR device which protects them from the high transmitter power. The TR cells takes a short time to break down completely after the transmitter pulse has started and a “spike” of energy leaks to the mixer diode during this period. If the power in the spike is too high the metal-to-semiconductor contact may overheat and burn out. Over a period of time even comparatively small leakage powers can cause the contact to change its rectifying properties and the diode becomes inefficient and ‘noisy’. This can be detected by a deterioration in the overall performance of the radar or by a decrease in the back-to-front resistance ratio of the diode.

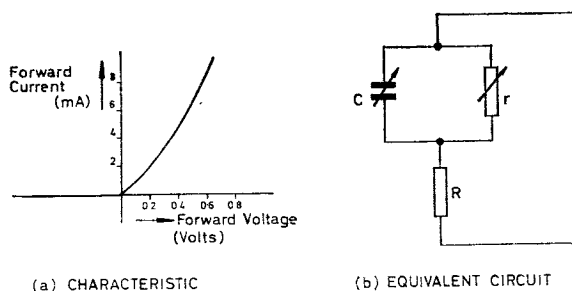


FIG 2. PROPERTIES OF A POINT-CONTACT DIODE

The Tunnel Diode

The tunnel diode is a semiconductor junction diode made of heavily doped germanium. It has a very narrow junction between the p- and n-type materials and current-carrying charges can “tunnel” through the junction at low values of forward bias. From 0 to about 50 mV the forward current due to these “tunnelling” carriers rises fairly sharply (region A in Fig. 3a) but as the forward voltage is increased above about 50 mV the tunnelling effect decreases and the forward current falls (region B). At about 350 mV normal semiconductor action takes over and the current again increases with an increase in forward voltage (region C).

The slope of the voltage/current curve in the region B is negative, that is, an increase in the forward bias voltage results in a decrease in the forward current and vice versa. By operating

the diode in this region it can be used as a negative resistance oscillator or amplifier. When employed in an oscillator circuit the positive resistance of the circuit is completely cancelled by the negative resistance of the diode; in an amplifier circuit the positive resistance is partly cancelled and an input applied to the circuit is amplified.

The equivalent circuit of a tunnel diode is similar to that of a

normal semiconductor diode (Fig. 3b) but the variable barrier resistance becomes negative ($-r$). The constant resistance (R) is due to the semiconductor material and the capacitance C depends upon the junction area and width. Thus C is fairly high (about 10 pF) and at microwave frequencies the tunnel diode presents a very low impedance.

The transit time of the carriers through the very narrow junction is very short (about 10^{-13} seconds) and so the tunnel diode can be used at the highest radar frequencies.

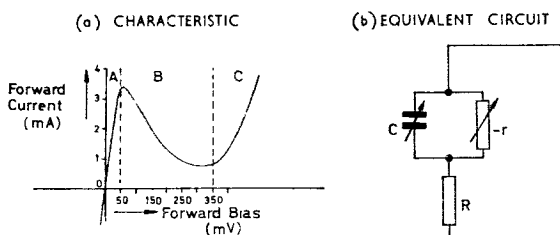


FIG 3. PROPERTIES OF THE TUNNEL DIODE

Microwave applications. One form of tunnel diode microwave oscillator is shown in Fig. 4a and its equivalent circuit in Fig. 4b. The tunnel diode is mounted at the end of a quarter wave length of microstrip which acts as a microwave tuned circuit. To bias the tunnel diode correctly a non-inductive slab of germanium is mounted at the end of the microstrip as shown. It has little

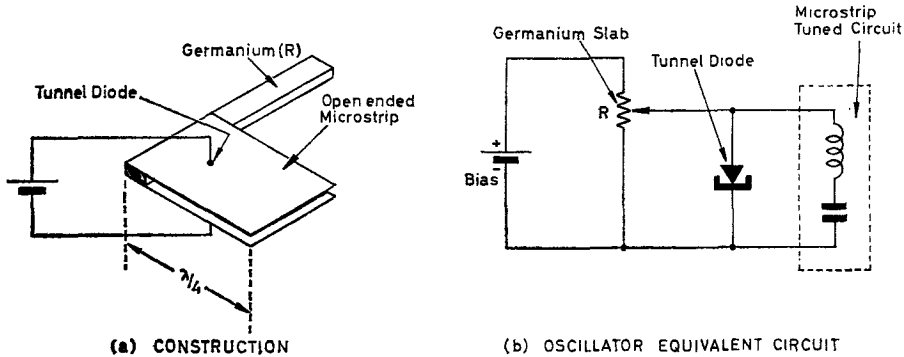


FIG. 4 TUNNEL DIODE OSCILLATOR

damping effect on the oscillations since it is at a voltage minimum. If the positive resistance of the circuit is less than the negative resistance introduced by the tunnel diode the circuit will oscillate. By varying the operating point, the negative resistance of the diode is changed and the output frequency can be varied.

The negative resistance of a tunnel diode may be adjusted so that the tuned circuit does not quite oscillate. A microwave input applied to the tuned circuit would then be amplified. The input and output share a common waveguide and a circulator must be used to separate the two as in Fig. 5.

Much development work is still being done on microwave applications of the tunnel diode. However, it has been successfully used as a low-power oscillator and amplifier at wavelengths down to the millimetre band. It has a reasonably low noise factor, medium bandwidth and good stability. The usual semiconductor advantages of compactness, reliability and simplicity make the tunnel diode a very promising microwave device.

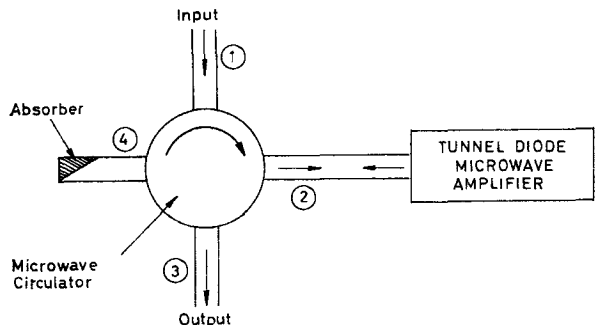


FIG. 5 TUNNEL DIODE AMPLIFIER

The P-I-N Diode

The p-i-n diode consists of a three-layer sandwich of semiconductor material. The p-i-n construction is formed by a thin wafer of undoped (*intrinsic*) silicon placed between a p-type and n-type layer of silicon. By varying the bias on a p-i-n diode its resistance to microwave energy is varied and the energy is either absorbed or reflected. In this way the output of a microwave oscillator may be amplitude modulated.

When a p-i-n diode is mounted in a waveguide or coaxial cable and *forward* bias is applied its resistance to microwave energy falls to almost zero. When *reverse* bias is applied its effective resistance is very high. Thus, when a p-i-n diode is mounted in a waveguide as shown in Fig 6a

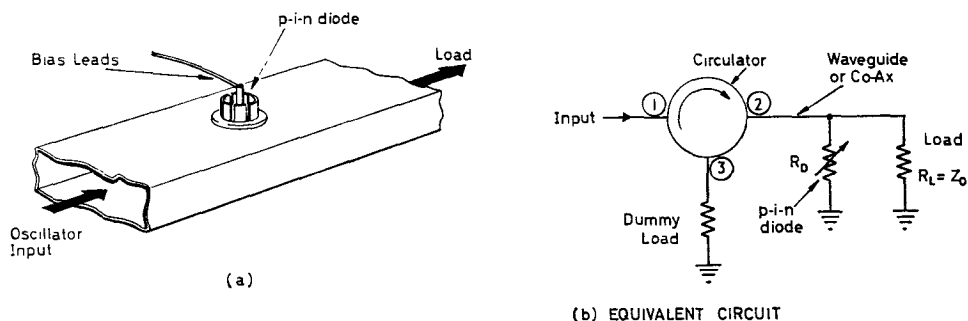


FIG 6 P-I-N DIODE MODULATOR

and is reverse biased it has little effect on microwave energy in the guide, most of which passes to the load. When the bias is switched to a forward value the diode acts as a short-circuit across the guide and input power is reflected back towards the circulator (Fig. 6b); little power reaches the load. Thus by switching the bias voltage to the p-i-n diode the oscillator output can be amplitude modulated.

Another application of the p-i-n diode eliminates the need for a keep-alive current in a TR cell (see p. 310). The diode is mounted across the cell and during the reception period it is in the open-circuit state (reverse bias or zero bias). Just prior to the transmitter firing, forward bias is applied so that the diode produces a short-circuit across the cell during transmission. This reduces the 'spike' leakage power and provides faster recovery time and longer life of the TR cell. The p-i-n diode requires much less power to operate than does the normal keep-alive system.

Advantages of the p-i-n diode when used at microwave frequencies are its small size and weight, fast switching speed, reliability and low switching power requirements. It can handle mean powers of up to 10W or peak powers of 500W and at present is used mainly in microwave test equipment.

The Variable Capacitance Diode (Varactor)

When a semiconductor p-n junction or point-contact diode is reverse biased a charge-free region is created in the vicinity of the junction. This is called the *depletion layer* and since it separates positive and negative charges on opposite sides of the junction, there exists across the junction a *capacitance*. If the reverse bias is *increased* the charges move further apart and the junction capacitance *decreases* because capacitance is inversely proportional to distance apart of the plates. If the reverse bias is *decreased* the depletion layer narrows and the junction capacitance *increases*. Thus by varying the value of reverse bias voltage applied to the diode the junction capacitance is varied.

Semiconductor diodes designed to make use of this variable capacitance are called *varactor diodes*. In the varactor diode the variation of capacitance with applied voltage is large compared with the minimum residual capacitance. The ohmic resistance (R) of the diode is small and over the operating range the variable barrier resistance in parallel with the capacitance is high and may be neglected. A high value of reverse breakdown voltage is desirable.

Thus the equivalent circuit of a varactor diode (Fig. 7a) consists of a variable capacitor (C) in series with a small constant resistance (R). Fig. 7b shows how the capacitance varies when the bias voltage is varied from a small forward value to a large reverse value.

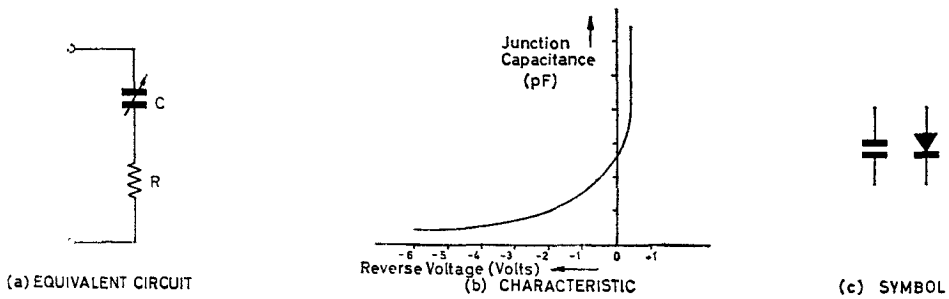


FIG. 7 CHARACTERISTICS OF THE VARACTOR DIODE

Point-contact germanium or gold-bonded diodes may be used as varactors for frequencies up to 6 Gc/s. Diffused junction diodes made of silicon or gallium arsenide can handle high powers. The diode is enclosed in a suitable case for mounting in strip-line or waveguide (Fig. 8).

The variable capacitance diode is used as a microwave switch, a frequency multiplier and as the variable capacitance in parametric devices.

Varactor microwave switch. The impedance of a varactor mounted across a waveguide as in Fig. 9a will depend upon the applied bias. Thus if a large reverse bias is applied the capacitance is small, its reactance to microwave energy in the guide is large and since it is effectively in parallel with the guide (Fig. 9b) it will have little effect on the energy, most of which will pass to the load.

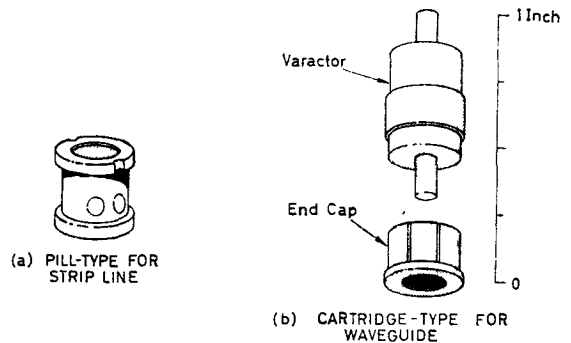


FIG. 8 TYPICAL VARACTOR DIODES

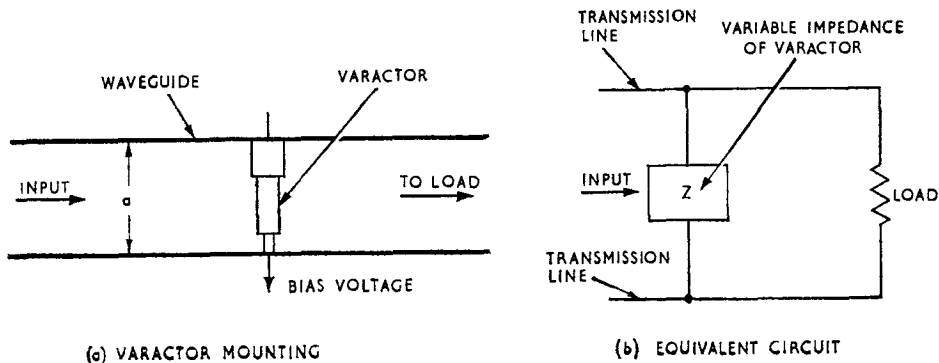


FIG. 9 VARACTOR AS A WAVEGUIDE SWITCH

When a small forward bias is applied the capacitance of the varactor is high, its reactance is low and it acts almost as a short-circuit across the guide. Energy travelling down the guide will be reflected and little will reach the load.

Varactor switches are small, require little switching power and have switching times measured in nanoseconds (10^{-9} seconds). Since they have small resistance they can handle high peak powers (up to 10kW) at frequencies up to 10 Gc/s.

Varactor frequency multiplier. The efficiency of a thermionic or semiconductor diode used to produce harmonics from a sine wave input is low because the diode acts as a non-linear resistor. A varactor acts as a non-linear capacitor and is almost loss-free. Thus high efficiencies (50 to 60 per cent) can be obtained by using the varactor as a microwave frequency doubler. As a frequency trebler its efficiency is somewhat lower.

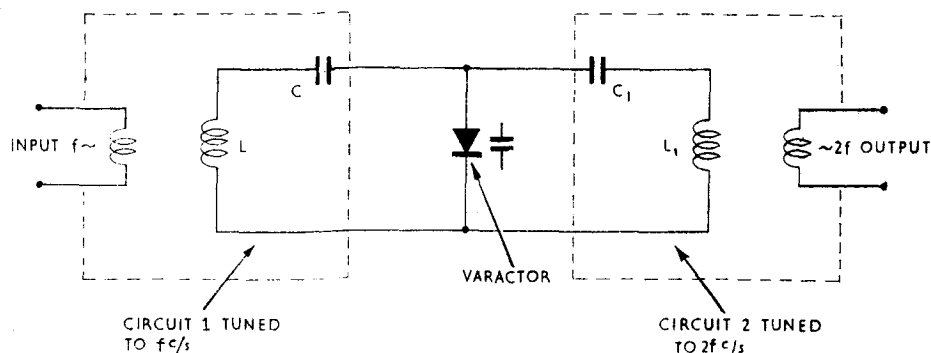


FIG. 10 BASIC FREQUENCY DOUBLER USING A VARACTOR

Fig. 10 shows the basic form of a varactor frequency doubler. Circuit 1 is tuned to the input fundamental frequency (f c/s) and develops a voltage at this frequency across the varactor. Circuit 2 is tuned to the second harmonic ($2f$ c/s) and provides the output.

A practical form of varactor frequency doubler used to convert an L band (1.5 Gc/s) input to an S band (3 Gc/s) output is shown in Fig. 11. The varactor is mounted across the waveguide

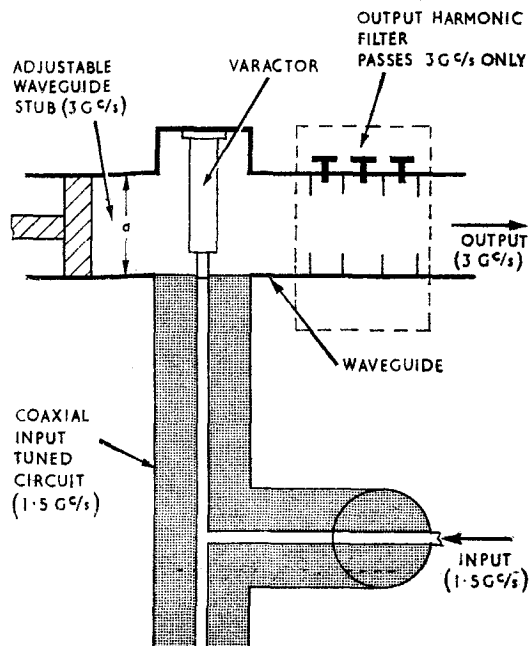


FIG. 11 PRACTICAL VARACTOR FREQUENCY DOUBLER

and is fed with the 1.5 Gc/s drive voltage via an input coaxial tuned circuit. A waveguide adjustable short-circuit stub is tuned to the second harmonic.

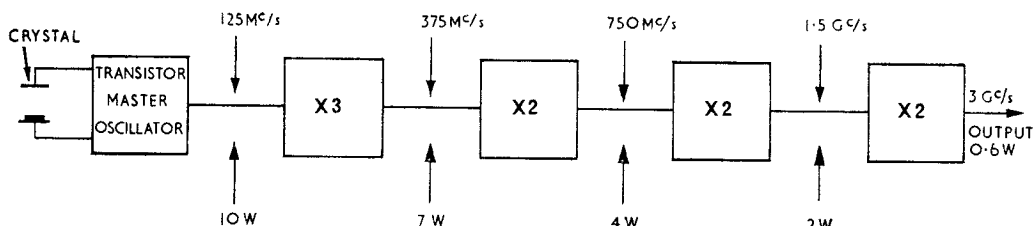


FIG. 12 TYPICAL SOLID-STATE FREQUENCY MULTIPLIER CHAIN

If the output from a crystal controlled transistor oscillator working at 125 Mc/s is fed into a chain of multiplier stages as shown in Fig. 12, an output at 3 Gc/s can be obtained.

This system is very compact, requires only small power supplies and has a low noise performance. Such a chain of varactor multipliers can be used as the local oscillator in microwave receivers and as the 'pump' source in parametric amplifiers (see later).

PARAMETRIC DEVICES

Introduction

In most conventional amplifying devices electrons transfer *d.c.* energy obtained from a power supply to *a.c.* energy at the required frequency. In a *parametric* device the value of a circuit *parameter* such as resistance, inductance or capacitance is varied and used to transfer energy from an *a.c.* source (e.g. an oscillator) to energy at the desired frequency.

One parametric device is the crystal mixer used in a superhet receiver (Fig. 13). Two *a.c.* inputs—the signal and local oscillator—use the variable *resistance* properties of the crystal to produce an output at the i.f. Since resistance is the variable parameter, attenuation—rather than amplification—of the input signal occurs.

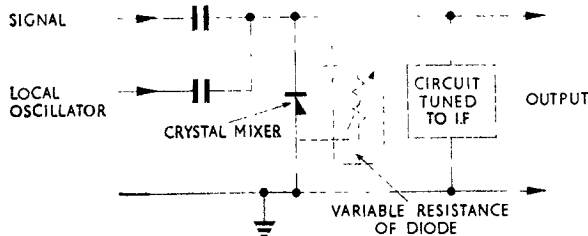


FIG. 13 VARIABLE RESISTANCE DIODE MIXER

Another parametric device is the saturable reactor (see Part 1B p. 340) in which *inductance* is the variable parameter (Fig. 14). The *d.c.* control signal input varies the inductance of an iron-cored coil and controls power from an *a.c.* source (the mains) to the load. The saturable reactor is an amplifier but because of high losses at r.f. it is normally used as a low-frequency or *d.c.* amplifier.

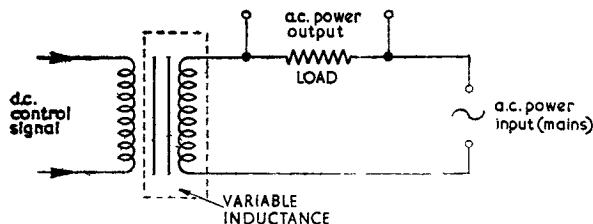


FIG. 14 VARIABLE INDUCTANCE AMPLIFIER (SATURABLE REACTOR)

frequency amplifier stage in a microwave radar receiver.

Microwave parametric devices use the low-loss variable *capacitance* properties of the varactor and have very low noise factors. Thus a parametric amplifier can be used as the signal

Principle of Parametric Amplification

Fig. 15a shows a tuned circuit LC which is considered loss-free and is energised in some way. The voltage across the capacitor is indicated by the solid line in Fig. 15b.

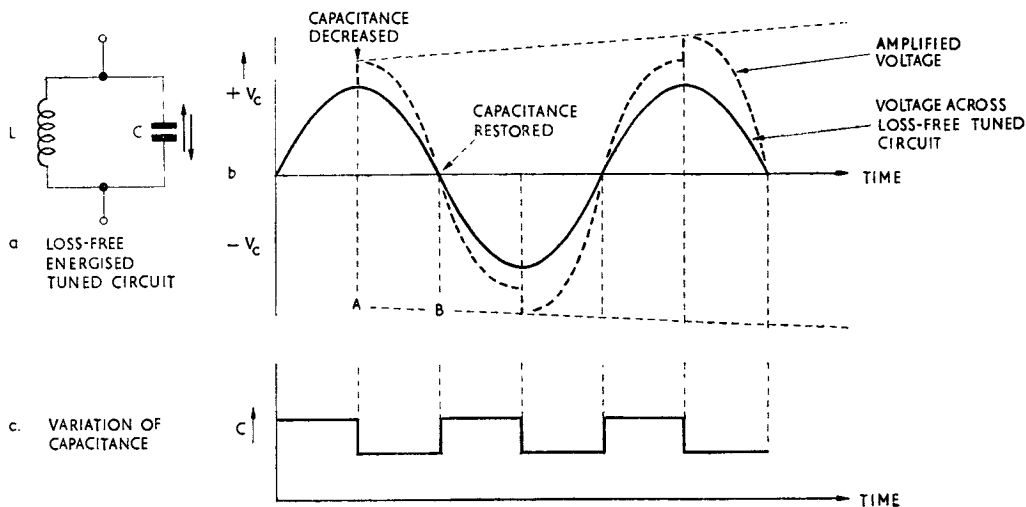


FIG. 15 PRINCIPLE OF PARAMETRIC AMPLIFICATION

Suppose that at instant A the plates of the capacitor are pulled further apart, i.e. the circuit capacitance is abruptly *decreased*. The charge Q cannot change and since the capacitance has fallen, the voltage across the capacitor must rise ($Q = CV$). This voltage rise is shown by the step in Fig. 15b. Thus the energy stored in the capacitor ($E = \frac{1}{2}CV^2$) has increased, the extra energy being provided by the work done in separating the plates against the force of electrostatic attraction.

At instant B the plates are returned to their original positions and since there is no voltage across the capacitor (and hence no stored energy), no energy is extracted from the circuit.

By 'pumping' the capacitor in this way there will be a continual transfer of energy from the pumping device via the variable capacitor to the circuit. The voltage across the capacitor and hence the energy stored in the tuned circuit will increase in steps as shown in Fig. 15b. Note that the variation of capacitance (Fig. 15c) must take place at exactly *twice* the circuit frequency.

In practice a signal is applied to a 'lossy' circuit and losses are made up from the pumping source. The 'pumping' device used is a varactor driven by an oscillator working at *twice* the signal frequency. This *pump oscillator* need not produce a square wave as in Fig 15c; a sine wave

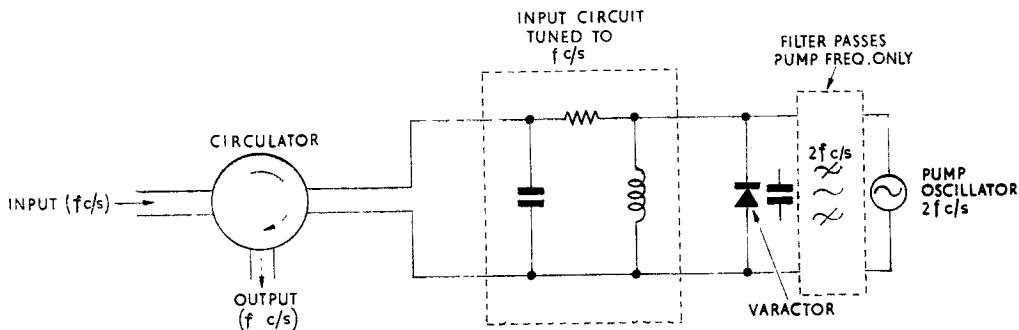


FIG. 16 ESSENTIALS OF A BASIC DEGENERATE PARAMETRIC AMPLIFIER

oscillator works just as well since most of the energy in a square wave is contained in the fundamental sine wave component. The amplified output is of the same frequency as the input and appears across the same two terminals. Therefore a circulator is necessary to separate input from output. A filter which passes the pump frequency only is placed between the varactor and pump oscillator. Thus a basic form of parametric amplifier would be as shown in Fig. 16.

In this simple form of parametric amplifier there are two frequencies; the input-output frequency (f c/s) and the pump frequency ($2 f$ c/s). Such an arrangement is called a *degenerate* parametric amplifier and for it to operate successfully an exact phase relationship between the two frequencies must be maintained. If, for example, the capacitance is *increased* instead of decreased at the peaks of capacitor voltage and returned to normal at the zeros, energy will be *extracted* from the circuit and the input is attenuated. At microwave frequencies this *phase coherence* between input and pump frequencies is impossible to attain.

Non-degenerate Parametric Amplifier

The need for this strict phase coherence can be overcome by introducing a third tuned circuit, called the *idler circuit*, into the amplifier (Fig. 17). The three circuits are coupled together by the varactor, and the frequency of the idler circuit (f_2 c/s) is chosen so that the pump frequency (f_p c/s) equals the sum of the input (f_1 c/s) and idler frequencies, i.e. $f_p = f_1 + f_2$.

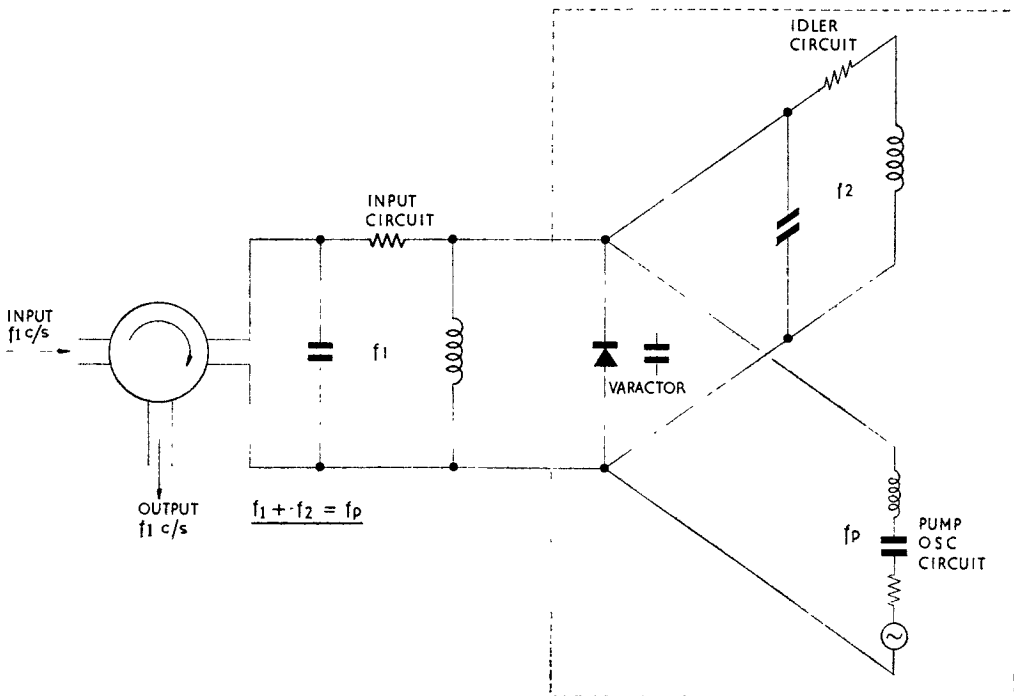


FIG. 17 BASIC NON-DEGENERATE PARAMETRIC AMPLIFIER

The input signal mixes with the pump frequency in the varactor and produces power at the idler frequency. Because of this mixing action, voltage in the idler circuit is always in the correct phase for amplification. The idler energy then combines with the pump energy to produce amplified signal power. Such a three-frequency device is called a *non-degenerate* parametric amplifier.

The pump circuit, idler circuit and varactor of Fig. 17 can be represented by a negative resistance, the effective value of which depends upon the pump frequency and power (Fig. 18).

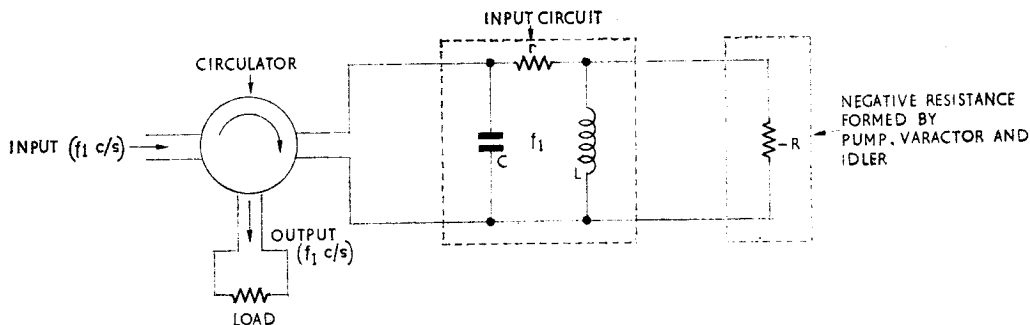


FIG. 18 EQUIVALENT CIRCUIT OF A NON-DEGENERATE PARAMETRIC AMPLIFIER

To obtain a high gain the negative resistance must be made *almost*, but not quite, equal to the positive resistance of the input circuit and load so that the total resistance is only *slightly* positive. A slight increase in pump power will cause the total resistance to become negative and the circuit will become an oscillator. This possible instability is a disadvantage inherent in all types of negative resistance amplifier.

Degenerate Parametric Amplifier

In the simple degenerate parametric amplifier only two frequencies (f_1 and f_p) are present; f_p must be exactly twice f_1 and phase coherence is essential. In practice, unless f_1 is *locked* to half f_p a third frequency, very close to f_1 , will be present. This third frequency adjusts itself to the correct phase for amplification and in fact acts as the idler frequency. It is sometimes called the *hidden idler*.

In this so-called degenerate parametric amplifier one tuned circuit supports both input and idler frequencies within its bandwidth. The pump frequency is adjusted as closely as possible to twice f_1 c/s.

Practical Microwave Parametric Amplifier

The tuned circuits used in microwave parametric amplifiers are resonant cavities, coaxial tuned circuits or resonant sections of waveguide. A reflex klystron or a chain of varactor multipliers can provide the pump power.

Fig. 19 shows a *non-degenerate* parametric amplifier employing a single resonant cavity in which a varactor is mounted. The pump power excites a mode with maximum E field at the varactor. The signal input, fed to the cavity through a circulator, sets up a different field pattern but still with maximum E field at the varactor. Energy at the idler frequency produces a third field pattern but again maximum E field occurs at the varactor. Amplified energy is extracted via the probe and fed to the load in the output arm of a circulator.

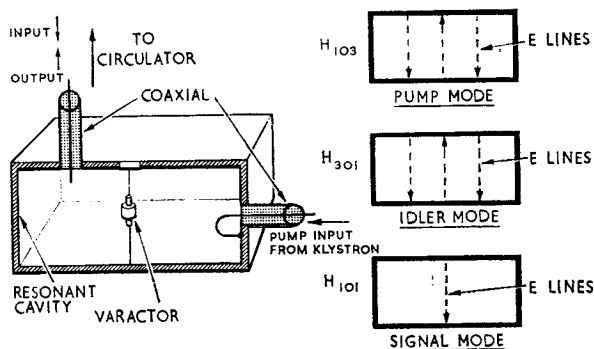


FIG. 19 RESONANT CAVITY NON-DEGENERATE PARAMETRIC AMPLIFIER

A *degenerate* parametric amplifier using a coaxial tuned circuit and waveguide is shown in Fig. 20. The bandwidth of the coaxial tuned circuit is sufficiently wide to support both input

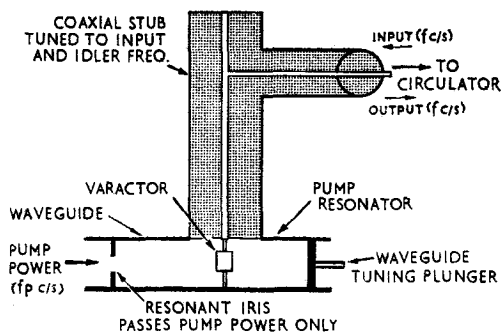


FIG. 20 DEGENERATE PARAMETRIC AMPLIFIER USING WAVEGUIDE AND COAXIAL

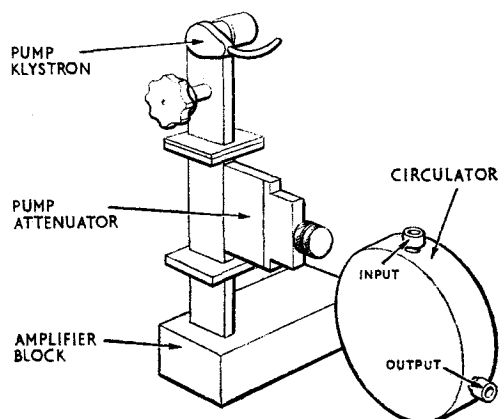


FIG. 21 1 Gc/s NON-DEGENERATE PARAMETRIC AMPLIFIER

and idler frequencies. Pump power at almost twice the input frequency is fed to the varactor through a waveguide. A circulator is again used to separate input from output.

The non-degenerate microwave parametric amplifier shown pictorially in Fig. 21 amplifies an input at 1 Gc/s. The pump klystron operates at 9.3 Gc/s and the idler frequency is 8.3 Gc/s. Its approximate size is 9 x 9 x 6 inches and it weighs about 11 lbs.

Properties of the Parametric Amplifier

The parametric amplifier has a very low noise factor, typically about 2 to 3 db, at room temperature. This is mainly because it does not employ a 'noisy' electron beam generated by a hot cathode. It is used as a signal frequency amplifier in a radar receiver to improve the receiver sensitivity and hence to increase the range of the radar. It is mounted between the TR cell and crystal mixer and is a 'fail safe' device, i.e. if the parametric amplifier ceases to operate the radar continues to function, although with a decrease in signal-to-noise ratio.

The noise present in a parametric amplifier is due to the noise introduced by the varactor and the thermal noise generated in leads, circulators, waveguide walls, etc. The noise factor increases as the operating frequency increases. If refrigeration is used to keep the amplifier at a temperature near 0° K even lower noise factors are obtained.

The gain of a parametric amplifier, which depends upon the pump power supplied and on the pump frequency, is limited by possible instability. Thus the output from a klystron pump oscillator is often fed to the varactor through an attenuator which is adjusted to give the required gain consistent with stability. Pump power is normally a few milliwatts.

For a given input frequency a high gain can be obtained by using a high pump frequency. However it is difficult to generate useful power at very short wavelengths. As shown in Table 1 a typical gain is 20 db and to obtain this with an input frequency of 600 Mc/s 20 mW of pump power at a frequency of 8.9 Gc/s is used.

Type	Input Frequency (Gc/s)	Idler Frequency (Gc/s)	Pump Frequency (Gc/s)	Band-Width (Mc/s)	Power Gain (db)	Noise Factor (db)
Non-degenerate ..	3	6.2	9.2	70	20	3
Non-degenerate ..	0.6	8.3	8.9	8	20	1
Degenerate	3	~3	~6	40	16.5	2

TABLE 1. DETAILS OF TYPICAL PARAMETRIC AMPLIFIERS

Because high-Q cavities, coaxial tuned circuits or waveguides are used, the bandwidth of a parametric amplifier is limited. The bandwidth depends upon the operating frequency and also upon the gain. As shown in Fig. 22 a decrease in gain results in a wider bandwidth.

The parametric amplifier can be tuned over a limited range by varying the pump frequency or by applying a d.c. bias to the varactor.

Other Parametric Devices

The negative resistance parametric amplifier is one of several parametric devices used in radar. In the non-degenerate amplifier the output is taken from the input circuit and is therefore at the same frequency as the input. If an output is taken from the *idler* circuit it will be at a different frequency from the input and the device becomes a *frequency converter*.

When the relationship between the three frequencies is the same as in the amplifier case ($f_p = f_1 + f_2$) the pump frequency (f_p) must always be greater than the output frequency (f_2) taken from the idler circuit. However, the idler circuit can be tuned so that it is either *above* or *below* the input circuit frequency. Thus we have either an *up-converter*, where f_2 is greater than f_1 , or a *down-converter* with f_1 greater than f_2 (Fig. 23). An increase in pump power results in an increase in conversion gain and like the negative resistance amplifier the circuit can become unstable.

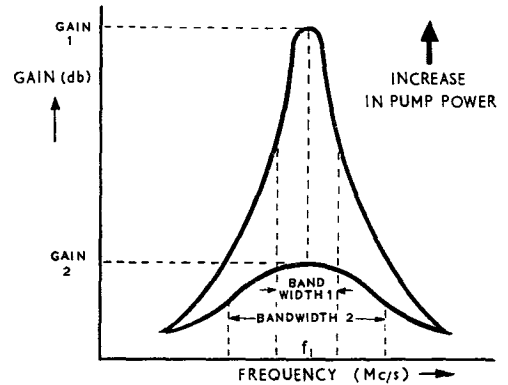


FIG. 22 GAIN/BANDWIDTH OF A PARAMETRIC AMPLIFIER

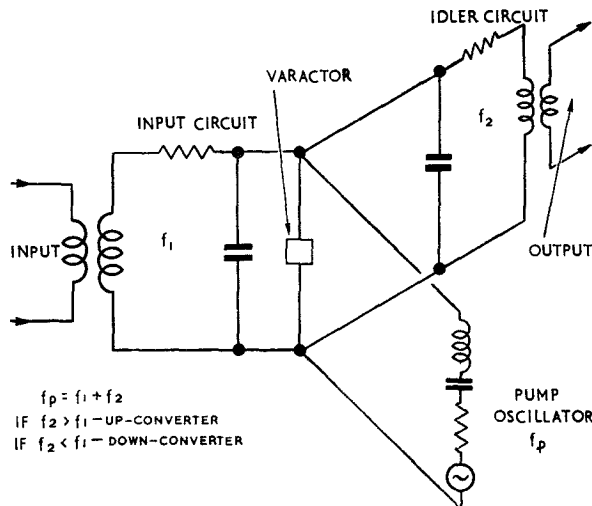


FIG. 23 BASIC CIRCUIT OF A NEGATIVE RESISTANCE PARAMETRIC FREQUENCY CONVERTER

A parametric frequency converter may be designed with a pump frequency *lower* than the *output* (idler) frequency. When the frequency relationship is $f_2 = f_p + f_1$ the device is called a parametric *up-converter* (*upper sideband*) to distinguish it from the negative resistance up-converter where $f_2 = f_p - f_1$. The conversion gain in this case is proportional to f_2/f_1 .

When the pump frequency is *lower* than the input frequency and the frequency relationship is $f_2 = f_1 - f_p$ the circuit acts as a *down-converter* since f_2 must always be lower than f_1 . In this case the pump circuit and varactor act as a frequency-changing coupler and there is no amplification. The device may be likened to a superhet mixing circuit with a variable reactance instead of the usual variable resistance of the mixer diode.

General details of the three-frequency parametric devices considered in this chapter are summarised in Table 2.

Name of device	Input freq	Output freq	Pump freq (f_p)	Idler freq (f_2)	Gain	Remarks
Negative resistance parametric amplifier	f_1	f_1	$f_1 + f_2$	$f_p - f_1$	Can be infinite	Degenerate amplifier if $f_1 \sim f_2$
Parametric oscillator	—	f_1	$f_1 + f_2$	$f_p - f_1$	Infinite	Low noise
Negative resistance parametric converter	f_1	f_2	$f_1 + f_2$	$f_p - f_1$	Can be infinite	If $f_2 > f_1$ up-converter If $f_2 < f_1$ down-converter
Parametric up-converter (upper-sideband)	f_1	f_2	$f_2 - f_1$	$f_p + f_1$	Proportional to f_2/f_1 — greater than unity	Positive resistance converter
Parametric down-converter	f_1	f_2	$f_1 - f_2$	$f_1 - f_p$	Proportional to f_2/f_1 — less than unity	Lower losses than diode mixer

TABLE 2. THREE-FREQUENCY PARAMETRIC DEVICES

CHAPTER 10

MASERS

Introduction

In previous chapters of this section we have seen the importance of a low noise factor in microwave amplifiers. The signal-to-noise ratio of a radar receiver sets a limit to the maximum range of the radar and the trend has been towards microwave signal amplifiers which improve this ratio. The low-noise travelling-wave tube (noise factor 2-3 db) and the parametric amplifier (noise factor of 1-2 db) give a considerable improvement in this direction (Fig. 1).

The *maser* is another type of microwave amplifier which has a very low noise factor. Its gain is similar to that of a parametric amplifier. At present it is used as a very low-noise signal frequency amplifier in ground equipments designed for satellite communication and radio astronomy.

In this chapter we shall consider the principles of maser amplification and the details of a practical maser.

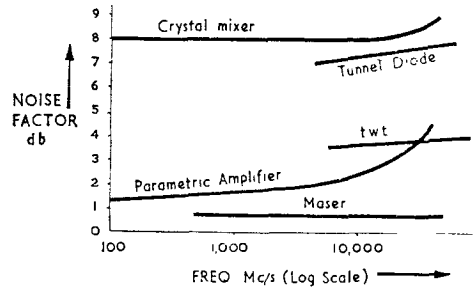


FIG. 1 NOISE COMPARISON OF MICROWAVE DEVICES

Maser Principles

All substances have atoms in which electrons spin in orbits round a nucleus, and an electron

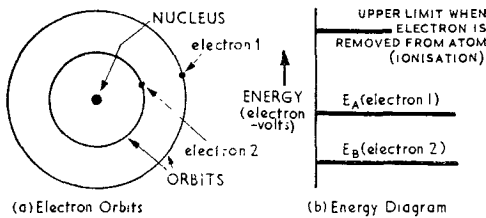


FIG. 2 ELECTRON ENERGY LEVELS

in an orbit has a value of potential energy due to its distance from the nucleus (Fig. 2a). The further an electron is from the nucleus the higher is its potential energy, i.e. electron 1 has greater potential energy than electron 2. Electrons can only occupy orbits of certain radii and their potential energy can be represented on an *energy diagram* (Fig. 2b). This is like a flight of uneven stairs, the electron energy being measured in electron volts ($1 \text{ eV} = 1.6 \times 10^{-19} \text{ joules}$).

Energy levels contain several sub-levels due to slight variations in the electron orbit and to energy of the electron spinning about its own axis. Fig. 3 shows these composite energy levels for E_A .

If a material is dense its atoms interact giving a large number of overlapping *energy bands*.

This type of material is of no use in masers, which use either a doped crystal (e.g. a ruby, where the maser-acting ions are scattered throughout a crystal lattice) or a gas.

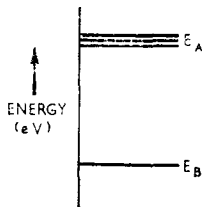


FIG. 3 COMPOSITE ENERGY LEVELS

Electrons falling from E_A to E_B emit radiation of frequency f such that $hf = E_A - E_B$ where $h = \text{Planck's constant} = 6.6 \times 10^{-34} \text{ joule-seconds}$. Electrons rising from E_B to E_A absorb radiation of the same frequency. Often, frequencies are in the visible light, ultra violet or even X-ray range but transitions between sub-levels of E_A (see Fig. 3) may, if proper choice is made, give rise to frequencies in the microwave region. This is done for masers.

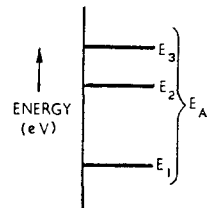


FIG. 4 ENLARGED ENERGY DIAGRAM OF FIG. 3

Enlarging the energy level for E_A , for example, we get the energy diagram of Fig. 4. In order to achieve emission, electrons must fall from E_3 to E_2 (or E_1), or from E_2 to E_1 and this means that there must be more electrons in the upper energy levels.

In thermal equilibrium the 'populations' of these energy levels are such that there are more electrons in the lower levels. The normal population state is as shown in Fig. 5. For the energies involved in masers the difference in populations is only a few per cent and is greater at low temperatures, e.g.

$$f_{12} = 10 \text{ Gc/s}, \quad \frac{N_1}{N_2} = 1.13 \text{ at } T = 4^\circ\text{K}.$$

At room temperature (300°K) the ratio would be very much lower.

If some electrons get to a higher level by absorbing energy they can fall to a lower state (not necessarily the original state) and in doing so they emit radiation. This occurs when energy is absorbed by heat, atomic collisions in electrical discharge tubes, fluorescence, etc. This emission is *spontaneous* and is characterised by the fact that upper levels do not become overpopulated.

In maser (and laser) action the upper level is arranged to be *overpopulated*, i.e. to have more electrons than the lower level, giving *population inversion*. A little spontaneous emission then occurs producing huge *stimulated emission*. This action can be compared to cumulative action in a multivibrator circuit. Stimulated emission cannot occur unless there is population inversion. Hence the name MASER which is made up of the initial letters of the phrase *Microwave Amplification by Stimulated Emission of Radiation*.

Overpopulation is produced by pumping. In the three-level maser—the only type we shall consider here—energy is supplied at the correct frequency to cause electrons to jump from energy state E_1 to E_3 so that $N_3 = N_1$ (Fig. 6). Then there are more electrons in state 3 than in state 2, (N_3 is greater than N_2) i.e. overpopulation of the higher energy state has been achieved. Electrons can now fall from E_3 to E_2 giving stimulated emission at frequency $f_{32} = \frac{E_3 - E_2}{h}$ c/s.

The build-up of overpopulation depends on how long electrons will stay in E_3 before falling to E_2 . This is known as the *relaxation time* and it must be long enough to allow overpopulation to occur, say 10^{-3} seconds. Spontaneous emission occurs after 10^{-8} seconds. The energy states with long lifetimes are rather rare, hence only a few materials are suitable for maser action. The best known is artificial *ruby* which has chromium ions scattered thinly throughout aluminium oxide. Low temperatures also help to give longer relaxation times.

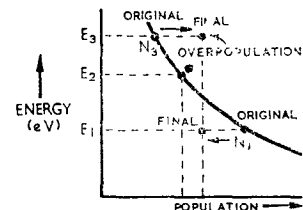


FIG. 6 OVERPOPULATION

Three-level Maser

When a *paramagnetic* substance is placed in a magnetic field its spinning electrons align themselves parallel or anti-parallel with the field. These alignments alter the energy level slightly and the two possibilities produce slightly different energies (Fig. 7). The difference between the two levels is proportional to the strength of the magnetic field.

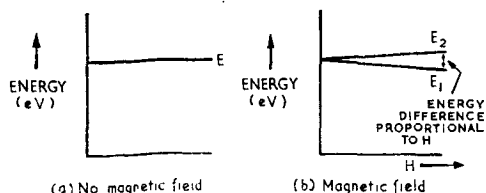


FIG. 7 EFFECT OF H FIELD ON PARAMAGNETIC SUBSTANCE

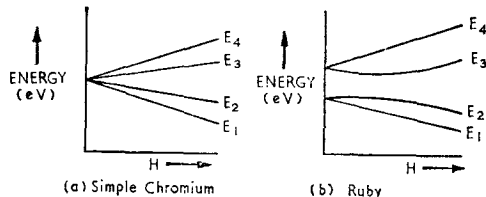


FIG. 8 VARIATION OF ENERGY LEVELS WITH STEADY MAGNETIC FIELD

In chromium there are three different spinning electrons which produce four energy levels between them when put in a magnetic field. Also, when the chromium is in ruby, other effects cause the energy diagram to be modified, especially when the magnetic field is not parallel to the crystal axis (Fig. 8). Certain energy level transitions do not occur and in maser design these diagrams have to be studied to find the allowed transitions to give the frequency desired by using a magnetic field of reasonable strength.

The magnetic field arrangement produces energy levels whose difference falls in the microwave, not optical range, and control of the H field gives control of frequency.

Three-level Cavity Maser

The general arrangement of a three-level maser using a resonant cavity is shown in Fig. 9. The paramagnetic ruby crystal is contained in a cavity which can resonate (in different modes) at the input signal frequency f_{32} c/s and at the pump frequency (f_{31} c/s). The cavity and crystal are immersed in a low-temperature bath of liquid helium and liquid nitrogen as shown. A strong steady polarising field splits the basic energy level of the crystal into separate levels necessary for maser action. By varying the strength of this field the energy spacings and thus the input and pump frequencies can be tuned within the bandwidth of the cavity.

Pump power at frequency f_{31} c/s is fed directly into the cavity to give reversal of populations between energy levels 3 and 1. The input signal to be amplified is applied via arm 1 of a circulator to the cavity and after being amplified by stimulated emission, travels via arm 3 to the next amplifying stage.

The paramagnetic ruby crystal when operated at 1.5°K has a relaxation time of about 5×10^{-3} seconds. A further advantage of operating the maser at liquid helium temperature is that noise due to thermal agitation in the crystal is reduced to a minimum. Like the parametric amplifier, the three-level maser uses a pump frequency higher than that of the input signal frequency and thus little noise is introduced by the pump source. Another similarity to the parametric amplifier is that the gain of the cavity maser can be increased by increasing the pump power and the maser can become unstable, i.e. it may oscillate.

A typical three-level cavity maser operating at 1.5°K amplifies a signal of 2.8 Gc/s with a gain of 20 db. The pump power is about 10mW at 9.4 Gc/s, the bandwidth 20 Mc/s and noise factor about 0.3 db. An X band maser amplifying a signal of 9.4 Gc/s requires a pump frequency of 24 Gc/s.

The disadvantages of a cavity maser are its tendency to oscillate (it is a negative resistance amplifier) and its narrow bandwidth. This latter is mainly due to the high-Q cavity used to magnify the weak input signal. Both these disadvantages are overcome in the *travelling-wave* maser.

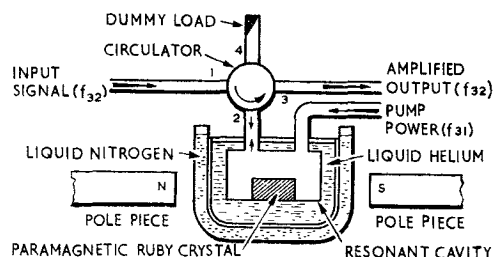


FIG. 9 OUTLINE OF THREE-LEVEL CAVITY MASER

The Travelling-wave Maser

The disadvantage of narrow bandwidth in a cavity maser can be overcome in the *travelling-wave* maser. Here, as in the travelling-wave tube amplifier, a slow wave structure is used to allow a weak input signal to interact with each part of the crystal for a longer period.

A type of slow wave structure often used in travelling-wave masers is the 'comb' structure shown in Fig. 10. This consists of wire 'fingers' projecting from the narrow wall of a waveguide.

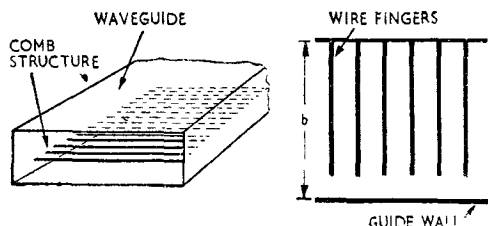


FIG. 10 'COMB' SLOW WAVE STRUCTURE

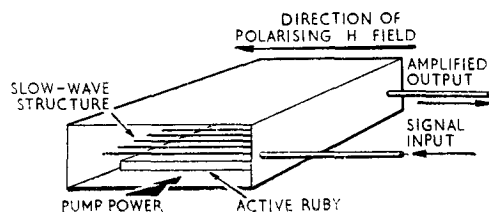


FIG. 11 BASIC PRINCIPLE OF A TRAVELLING-WAVE MASER

Each finger, and that part of the waveguide below it, form a transmission line along which the wave travels as it winds down the guide.

The active ruby is mounted *below* the slow wave structure as shown in Fig. 11. The input signal is applied via a coaxial lead and the output is taken from the end of the slow wave structure. Pump power is fed down the waveguide causing population inversion in the crystal. The input signal causes stimulated emission and is thus amplified. A magnetic field in the direction shown is provided and the whole structure is immersed in a liquid helium bath.

The stability of the travelling-wave maser is improved by placing a crystal of *heavily doped* ruby *above* the wire fingers (Fig. 12). This 'dark' ruby attenuates energy travelling in the reverse direction to the signal thus reducing reflections and preventing oscillation.

By increasing the length of active ruby the gain of the travelling-wave maser can be increased. It can be tuned over a wide bandwidth by altering the strength of the polarising magnetic field.

A typical travelling-wave maser operating at 19 Gc/s has a gain of 20 db, a bandwidth of 55 Mc/s and a noise factor of 0.16 db.

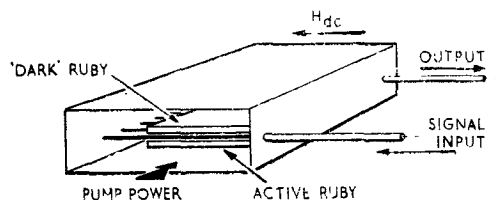


FIG. 12 THE TRAVELLING-WAVE MASER

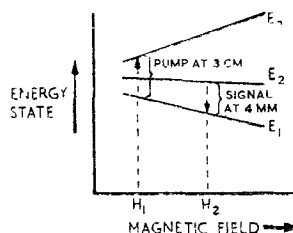


FIG. 13 VARIATION OF ENERGY SPACINGS WITH POLARISING FIELD

Millimetre Wave Masers

In the three-level solid-state masers discussed so far the pump frequency must be higher than the signal frequency. Thus to operate a three-level maser at millimetre wavelengths we have the difficulty of generating pump power at the very short wavelengths involved.

One way of overcoming this difficulty is to use a substance in which the energy levels diverge as the polarising field is increased (Fig. 13). With a fairly low value field (H_1), a comparatively low frequency (X band) pump raises E_1 electrons to E_3 level thus creating a population inversion between E_1 and E_2 levels. The H field is then rapidly increased (within the relaxation time) to H_2 . Stimulated emission can now occur at a frequency f_{21} c/s corresponding to the new spacing between E_2 and E_1 .

High values of magnetic polarising field are required (2.5 Wb/m^2 at 4mm) and it is difficult to rapidly switch such a high field.

Lasers

Considerable development work has been done in recent years on the LASER (*Light Amplification by Stimulated Emission of Radiation*). Like masers they work by overpopulation but the transitions are between main energy levels. Thus energy differences and hence frequencies are higher. No magnetic field is required for laser operation.

The main types of laser are gas (a mixture of helium and neon) and solid (ruby) but there are also semiconductor lasers which work on a different principle.

SECTION 5

PULSED RADAR TRANSMITTERS

Chapter		Page
1	Outline of Pulsed Radar Transmitters	361
2	Modulators	373
3	Transmitter Power Supplies	383

CHAPTER 1

OUTLINE OF PULSED RADAR TRANSMITTERS

Introduction

The purpose of a pulsed radar transmitter is to generate equally-spaced high-power r.f. pulses of short duration. The pulses must be well-shaped, with steep leading and trailing edges, for accurate range measurement and good target discrimination. Although radars operate over a wide range of frequencies and power, and vary considerably in size and design, a pulsed radar transmitter consists of three basic parts as shown in Fig. 1.

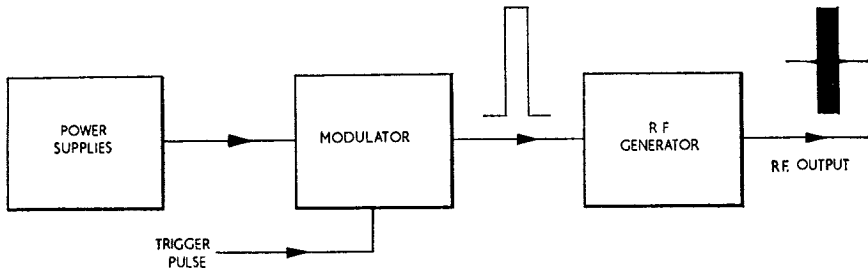


FIG 1. MAIN PARTS OF A PULSED RADAR TRANSMITTER

The r.f. generator. This is the source of r.f. power. The type of r.f. generator used in the transmitter depends mainly on the frequency. At centimetric wavelengths it may be a high-power multi-cavity klystron, a travelling-wave tube or a magnetron. At u.h.f., disc-seal valves or multi-cavity klystrons can be used, while at v.h.f. and lower frequencies conventional valves are suitable.

The modulator. This produces high-power pulses at the radar p.r.f. which are used to modulate the r.f. generator. Because of the different characteristics of various radars the modulator may take several different forms. Sometimes it generates the master timing pulse with which other circuits in the radar are synchronised.

Power supplies. The transmitter requires a number of different power supplies. These include h.t., l.t. and bias supplies for the valves and an e.h.t. supply of several thousand volts to operate the modulator. Stabilisation circuits are used to ensure a non-fluctuating supply.

In this section we shall consider the modulator and power supply circuits and see how these are used in conjunction with the r.f. generator to provide the transmitter output.

Transmitter Range

We already know that the power received through a given area is inversely proportional to the *square* of the distance from the source. Thus to *double* the range of a transmitter we would have to increase the power output by a factor of 4. In radar we are concerned with the strength of a returned echo and thus to maintain the same strength echo at the receiver from a target at *twice* the range, the peak power of the transmitter must again be raised by a factor of 4. Thus to double the range of a pulsed radar it is necessary to increase the peak power by a factor of 16, i.e. 2^4 . Similarly, to treble the range the power must be raised by 3^4 or 81. If the power is reduced by half, range will be reduced to 84% of its former value (*see* Fig. 2). However, the peak power of a radar transmitter does not alone decide the maximum range. Other factors have to be considered; the more directional the transmitter/receiver aerial the greater is the power

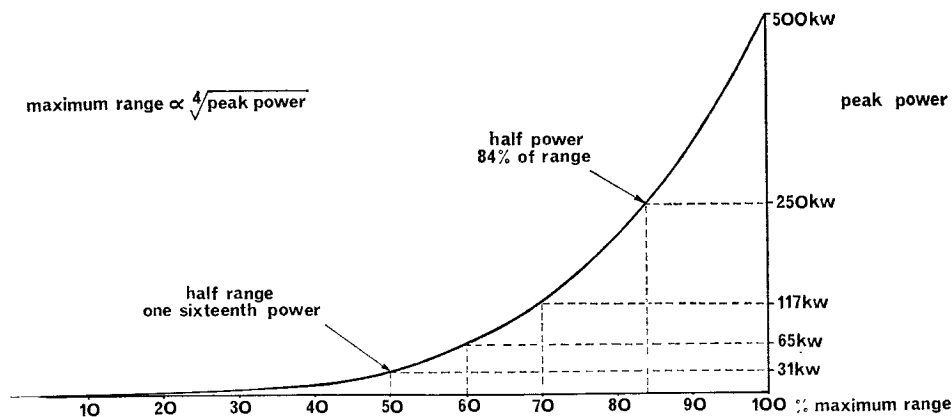


FIG 2. RELATION BETWEEN PEAK POWER AND RADAR RANGE

directed to and received from a target; the larger the radar cross-section of the target the stronger is the returned echo. The minimum energy detectable by the receiver also governs the maximum range. This last factor means that a pulse of long duration but small peak power may be more easily detected than a short pulse of high peak power. It is the pulse *energy* (power $\times$ time) that is important. These factors are discussed in Section 6.

Pulse Duration and PRF

To detect a target at close range the radar should radiate a short-duration r.f. pulse. Further, with a short-duration pulse good range discrimination between nearby targets is possible. Thus for an airborne interception (AI) radar a pulse duration of about $0.2\mu\text{s}$ is required, while for a long-range ground search radar much longer r.f. pulses (up to $10\mu\text{s}$) can be used.

The pulse repetition frequency (p.r.f.) is the number of r.f. pulses transmitted per second and is related to the maximum range of the radar, because the echo from one pulse must return to the radar before the next pulse is transmitted or false target ranges will be displayed. When a narrow beam is swept at a fast rotational speed a high p.r.f. is desirable so that a reasonable number of returns per sweep is obtained.

Mean Power

For a given transmitter, the mean r.f. power output available is determined by the type of transmitter valve, power supplies, etc. and is equal to:

$$\text{Peak power} \times \text{pulse duration} \times \text{p.r.f.}$$

Thus for a given mean power the peak power, pulse duration and p.r.f. can be adjusted to suit the requirements of the radar. For example, if the available mean power of a transmitter is 250 watts, three of the possible combinations are:

- Peak power = 250 kW, pulse duration = $1\mu\text{s}$, p.r.f. = 1000 p.p.s.
- or Peak power = 500 kW, pulse duration = $0.5\mu\text{s}$, p.r.f. = 1000 p.p.s.
- or Peak power = 100 kW, pulse duration = $2\mu\text{s}$, p.r.f. = 1250 p.p.s.

In some multi-range radars these three factors can be varied to suit each range.

Types of Radar Transmitter

The basic pulsed radar transmitter shown in Fig. 1 can take one of two forms.

The high-power oscillator. This is the most common type of radar transmitter in which a

high-power oscillator, such as a magnetron, is switched on and off by the modulator. The modulator may be a high-power stage as in Fig. 3a, producing a well-shaped e.h.t. pulse which directly modulates the high-power oscillator; or it may consist of several stages in which a low-power pulse from the master timer is shaped and amplified by a driver stage and pulse power amplifier stage before being applied to the r.f. oscillator (Fig. 3b). In both cases the r.f. energy of the pulse is generated in one stage (the oscillator) and forms the final r.f. output.

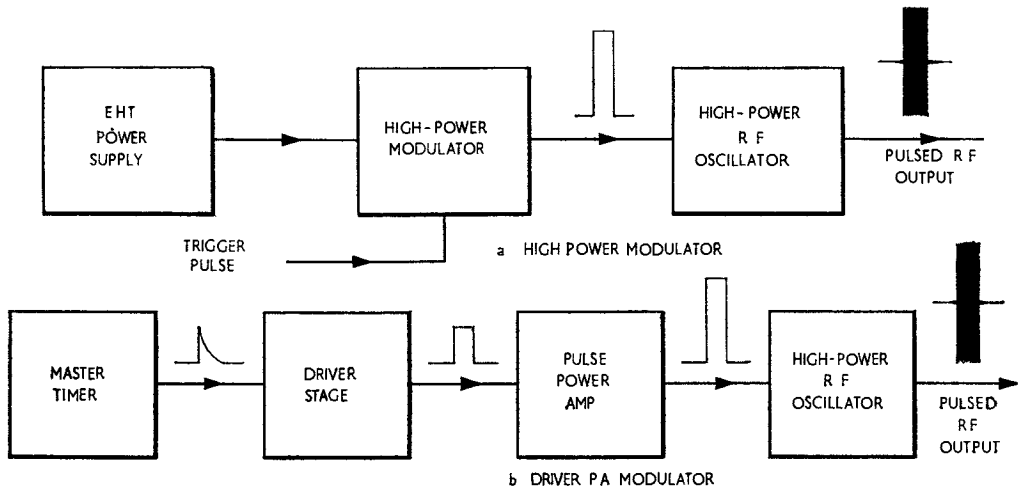


FIG. 3. HIGH-POWER OSCILLATOR TRANSMITTERS

The MOPA. In the *master oscillator-power amplifier* type of pulsed radar transmitter (Fig. 4) a low-power r.f. oscillator is amplitude modulated and its output is then amplified by an r.f. amplifier before being radiated. The master oscillator may be a crystal controlled oscillator followed (if necessary) by frequency multipliers, or it may be a stable-frequency resonant cavity oscillator. The r.f. power amplifier can be a multicavity klystron, a travelling wave-tube, an amplitron or conventional or disc-seal power valves depending on the output frequency.

The advantages of the m.o.p.a. over the power oscillator transmitter are; a much simpler modulator is required since it does not have to provide all the power for the output r.f. pulse; the frequency stability of the output is much better; and phase coherence between radiated pulses is much easier to obtain. This last advantage is important in MTI radars (Section 7).

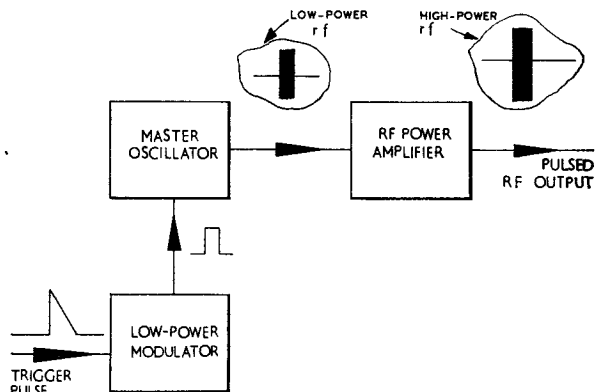


FIG. 4. MASTER OSCILLATOR—POWER AMPLIFIER TRANSMITTER

Transmitter Valves

In previous sections we have considered the construction and action of various types of valves suitable for use in radar transmitters and we shall now revise their main characteristics. The type of transmitter valve used depends mainly on the frequency and power output of the

transmitter. By far the most commonly used valve in centimetric radars is the *magnetron*. This is used as a high-power oscillator and when modulated by a suitable high-power modulator can provide well-shaped narrow pulses of r.f. with peak powers of several megawatts. It is cooled by cold air blast on cooling fins, or in the higher power types, by water cooling.

Centimetric *ampliftrons* may be used in the m.o.p.a. type of transmitter. They amplify the modulated low-power output from a stable r.f. oscillator and are capable of very high peak power outputs (up to 10 MW). They are broad-band devices and have high efficiencies.

Multi-cavity klystrons may also be used as centimetric high-power transmitter valves and like the ampliftron can be driven by a low-power oscillator in an m.o.p.a.-type transmitter.

Travelling-wave tubes are used as power amplifiers in some radars and can provide peak power of up to 100 MW. They amplify a low-power input and because of their wide bandwidth the output can be frequency modulated.

Magnetrons, ampliftrons, t.w.t.s and multi-cavity klystrons can be designed to operate at frequencies in the u.h.f. band (300-3,000 Mc/s). They are much larger than their centimetric counterparts and find use mainly in long-range ground radars.

Above 300 Mc/s the conventional cylindrical valve construction is not used because of lead inductance and capacitance and transit time effects. At frequencies up to 1,000 Mc/s grid control valves have a *planar* electrode construction suitable for use with coaxial line resonators and resonant cavities. These valves can be used as the p.a. stage in a pulsed radar transmitter and are capable of providing a mean power of 1 kW at 1,000 Mc/s. Multi-unit planar triodes and tetrodes, connected directly to resonant cavities have a mean power of 10 kW with a peak power of 2 MW. They are physically much smaller than klystrons and may be cooled by water or air blast.

Transmitter Oscillator Circuits

The form of tuned circuits used in radar transmitters depends to a large extent on frequency. Above about 200 Mc/s the conventional LC tuned circuit is impracticable and there is a choice of resonant cavities, concentric line resonators or lecher bars. Resonant cavities are generally used in centimetric radars where they are small and often form an integral part of a valve such as a magnetron or klystron. They are also used in conjunction with grid-controlled planar valves on ground u.h.f. radars where their large size is no disadvantage. Thus although generally considered as a microwave device the resonant cavity, owing to its low losses, finds considerable use at much longer wavelengths.

Concentric line resonators form suitable tuned circuits in the upper u.h.f. and lower centimetric ranges and are compact, easily tuned and have a high Q. The disc-seal valve electrodes fit directly on to the resonator and together, the valve and resonator, form a compact and convenient amplifier or oscillator.

Lecher bar resonators are used in radars operating between 200 and 400 Mc/s. They are short sections of open transmission line, usually made of silver-plated copper tubing to reduce skin effect losses. The line is adjusted to $\frac{1}{4}\lambda$ by a movable shorting bar at one end, when it acts as a parallel tuned circuit.

Ultra audion oscillator. This is a v.h.f. oscillator using a conventional triode and lecher bars as the tuned circuit. It is a modified Colpitts circuit in which the valve inter-electrode capacitances form part of the tuned circuit (Fig. 5a and b). In the practical circuit (Fig. 5c) the capacitor C acts as a short-circuit at the frequency of oscillations and by adjusting its position on the lecher bars until they are $\frac{1}{4}\lambda$ long, a parallel tuned circuit is formed by the lecher bars. Automatic bias is provided by C and R and the power supply leads are filtered to prevent r.f. reaching the power supplies.

Push-pull lecher bar oscillator. By using two valves connected in push-pull, a high-power output with high efficiency is obtained. The inter-electrode capacitances are in series across the

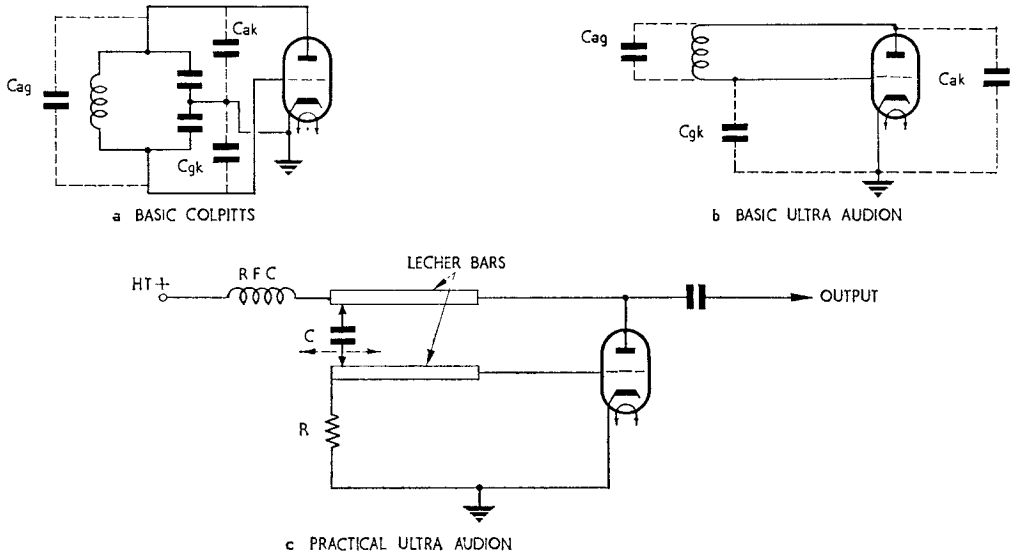


FIG. 5. THE ULTRA AUDION OSCILLATOR

tuned circuit reducing the total capacitance and thus enabling oscillation at higher frequencies to occur.

Fig. 6 shows the outline of a push-pull lecher bar oscillator suitable for a radar transmitter. The circuit acts as a t.a.t.g. oscillator, feedback being via the C_{ag} of the valves. Class B or C self bias is provided by $C_g R_g$ and the valves conduct alternately. Specially shaped valves are used for short lead connections to the lecher bars. An output match is provided by the sliding taps. All supply leads are filtered.

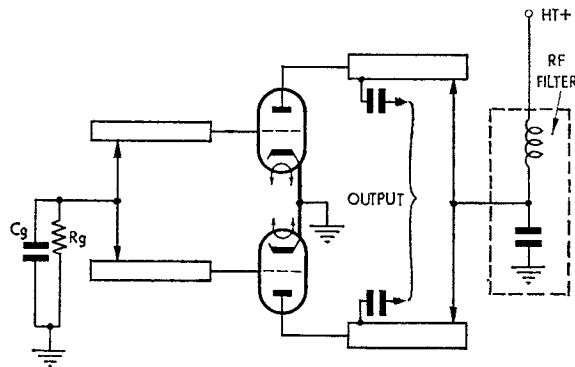


FIG. 6. PUSH-PULL LECHER BAR OSCILLATOR

Delay Lines

Delay lines are used in radar to impose an exact time delay on a pulse (Fig. 7a) or to form a narrow well-shaped pulse suitable for modulating a radar transmitter (Fig. 7b). In this latter role they are widely used in radar modulators and can be designed to produce very short pulses ($0.1 \mu s$) or longer pulses ($10 \mu s$). The amplitude of the pulse can be very large and depends upon the voltage applied to the delay line. If a high-voltage pulse of several kilovolts is formed it

can be used to modulate directly the transmitter oscillator. Lower amplitude pulses may be amplified by a pulse power amplifier before being applied to a high-power r.f. oscillator.

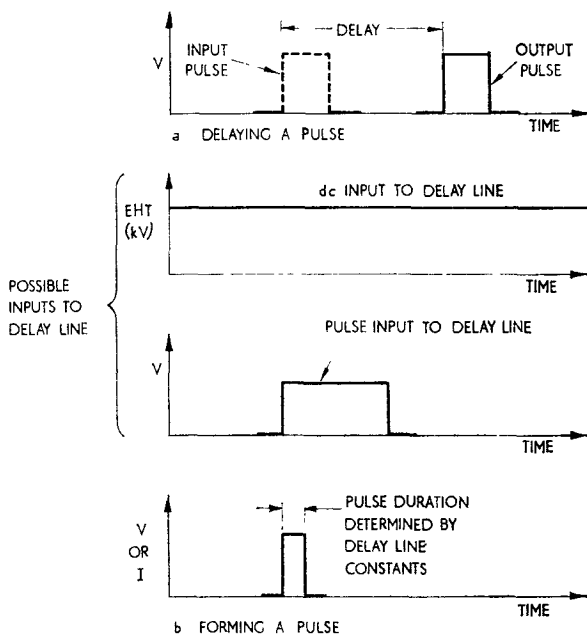


FIG. 7. USES OF DELAY LINES

In transmission line theory we learned that a loss-free open wire transmission line can be considered to be made up of an infinite number of sections each comprising inductance and capacitance as in Fig. 8b. With a uniform line the L and C of each section are equal. If a voltage step is applied to one end of the transmission line it will travel down the line at a velocity depending upon the values of L and C per section and will take a definite time to reach the other end. This velocity is given by $v = \frac{1}{\sqrt{LC}}$ sections per second where L and C are measured in henries and farads per section respectively.

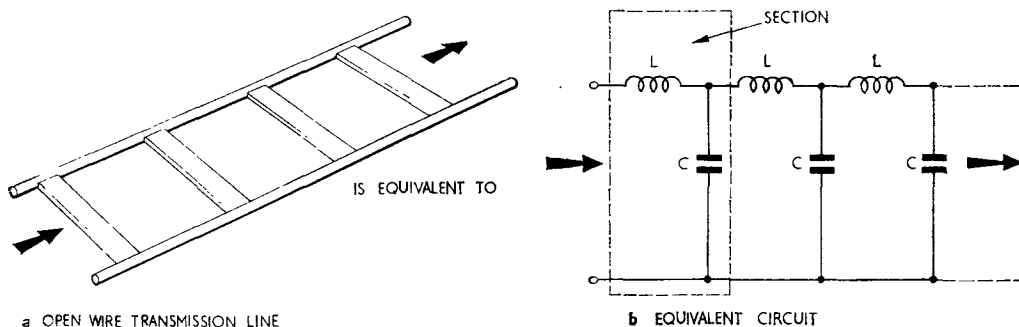


FIG. 8. TRANSMISSION LINE EQUIVALENT

In a normal open wire transmission line the velocity of propagation is not much less than that in free space because L and C are small. Thus to obtain a time delay of $1\mu s$ between its ends a line almost 300 metres long would be needed.

One form of delay line consists of actual capacitors and inductors wired together in sections as in Fig. 9. In this way a very compact structure can be made, providing delays of several

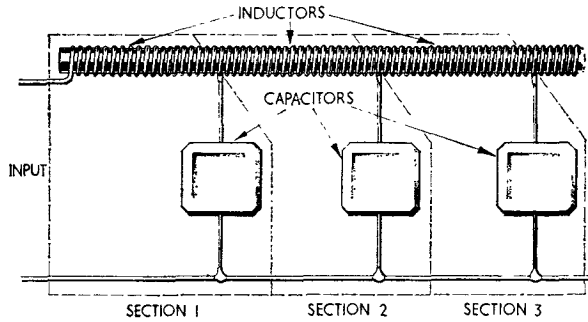


FIG. 9. DELAY LINE

microseconds. Because L and C are relatively large a delay line much shorter than 300 metres can be used to provide a delay of $1 \mu s$. The total delay is given by $n\sqrt{LC}$ seconds where n is the number of sections in the artificial line and L and C are the inductance and capacitance per section. The characteristic impedance (Z_0) of the line is given by $\sqrt{\frac{L}{C}}$ ohms, as with a transmission line.

Fig. 10a shows a delay line connected to a pulse generator of voltage E and source resistance (R_s) equal to Z_0 of the line, and terminated with a resistance equal to Z_0 . The leading edge of the pulse charges up the capacitors of each section in turn to $E/2$ volts, through the associated

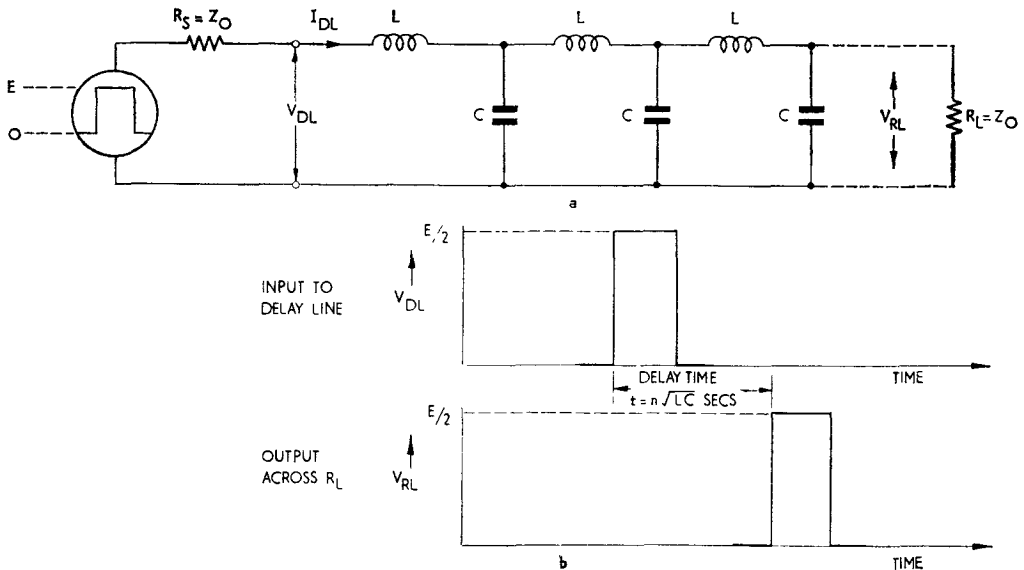


FIG. 10. DELAY LINE USED TO DELAY A PULSE

inductances and after a delay equal to $n\sqrt{LC}$ seconds this leading edge appears across R_L (Fig. 10b). If R_L equals Z_0 , no energy is reflected and the current (I_{DL}), flowing through R_L during the steady part of the input pulse, is constant.

When the input pulse voltage falls to zero each section of the delay line discharges in turn

maintaining the same current through R_L until, after $n\sqrt{LC}$ seconds, the last section discharges and the trailing edge appears across R_L .

There are other types of delay line which do not use a combination of inductance and capacitance. Their function is the same as that discussed here, i.e. they impose a time delay on the passage of a pulse through them. However, they are seldom used in modulator circuits and we shall learn more about them in Section 7.

Modulator Pulse Forming Networks

LC delay lines are used in radar modulators to form pulses of the required duration. The delay line with the necessary number of sections may be terminated in an open-circuit or a short-circuit and is first charged from the power supply and then discharged through the load. We shall now consider the principles involved.

Charging an Open-end Delay Line

Fig. 11a shows an open-end delay line connected to a d.c. source of V_s volts through a switch and charging resistor R_c , equal to the characteristic impedance (Z_0) of the line. When the switch is closed at time t_1 (Fig. 11b) the line appears to the voltage source as a resistance of value Z_0

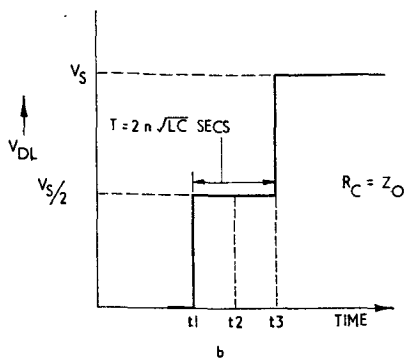
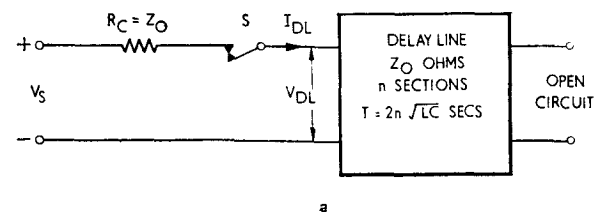


FIG. 11. CHARGING AN OPEN-END DELAY LINE

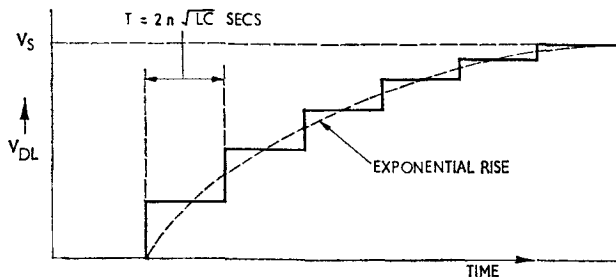


FIG. 12. CHARGING THROUGH A RESISTOR GREATER THAN Z_0

ohms. Thus initially, half V_s is developed across R_c and half across the delay line input terminals. This voltage wavefront travels down the line, accompanied by a current wavefront and each section of the line is in turn charged to $V_s/2$ volts. After $n\sqrt{LC}$ seconds (t_1 to t_2 in Fig. 11b) both wavefronts arrive at the open end and the current through the last inductor of the delay line falls to zero. This produces a back e.m.f. which charges the last capacitor to V_s volts. Thus the voltage is reflected *in phase* and the current in *antiphase* and both wavefronts move back along the line charging each section in turn to V_s volts as the current falls to zero. After a further $n\sqrt{LC}$ seconds (t_2 to t_3) the wavefronts arrive back at the delay line input terminals and the line is charged to V_s volts. Note that the time taken for the delay line to fully charge to V_s volts (t_1 to t_3) is $2n\sqrt{LC}$ seconds. Because R_c is equal to Z_0 there is no further reflection and the action stops.

The action described above is not confined to delay lines. Ordinary d.c. circuits act in the same way but the times involved are so short that they are imperceptible. In radar circuits, however, these short time intervals are all-important.

If the charging resistor R_c is *greater*

than the line Z_0 , V_{DL} will initially rise to a value *less* than $V_s/2$ volts. Reflections then occur when the wavefront returns to the source and V_{DL} rises in a series of steps each of T seconds duration (Fig. 12), and eventually reaches the value of the supply.

Discharging the Delay Line

To form a pulse, the delay line is discharged through a resistance R_L equal to Z_0 of the line, as in Fig. 13a. When the switch is closed the voltage across R_L rises to $V_s/2$ volts (Fig. 13b) and that across the delay line falls to $V_s/2$ volts. Thus a negative wavefront moves along the line discharging each section in turn to $V_s/2$ volts. At the open-circuited end the wavefront is reflected and travels back discharging each section to zero volts. Therefore a pulse of amplitude $V_s/2$ volts and duration $2n\sqrt{LC}$ seconds is developed across R_L .

In a radar modulator the switch S is normally a gas-filled valve such as a thyratron and the load resistance R_L is the transmitter impedance. If this impedance is not correctly matched to the characteristic impedance of the delay line the line discharges in a series of steps or pulses of decreasing amplitude. In most delay line modulators it is essential to prevent these conditions, i.e. the load must be correctly matched to the delay line, usually by means of a transformer.

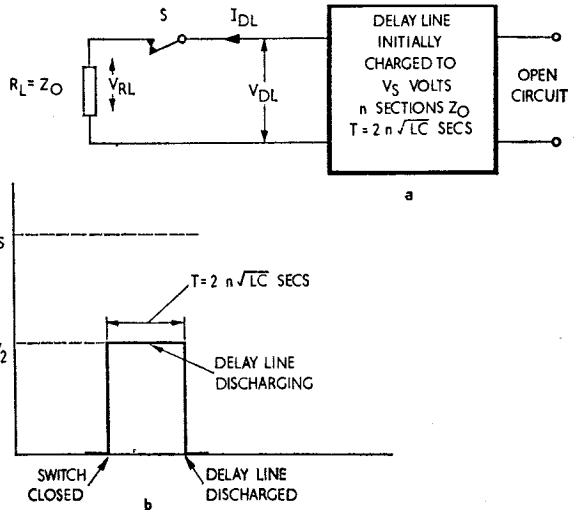


FIG 13. OPEN-CIRCUIT DELAY LINE DISCHARGE

Short-circuited Delay Line

Fig. 14a shows a delay line terminated in a *short-circuit* and charged from a d.c. supply of V_s volts, through a charging resistor R_C equal to the Z_0 of the delay line. Across the short-circuit termination the *voltage* must at all times be zero.

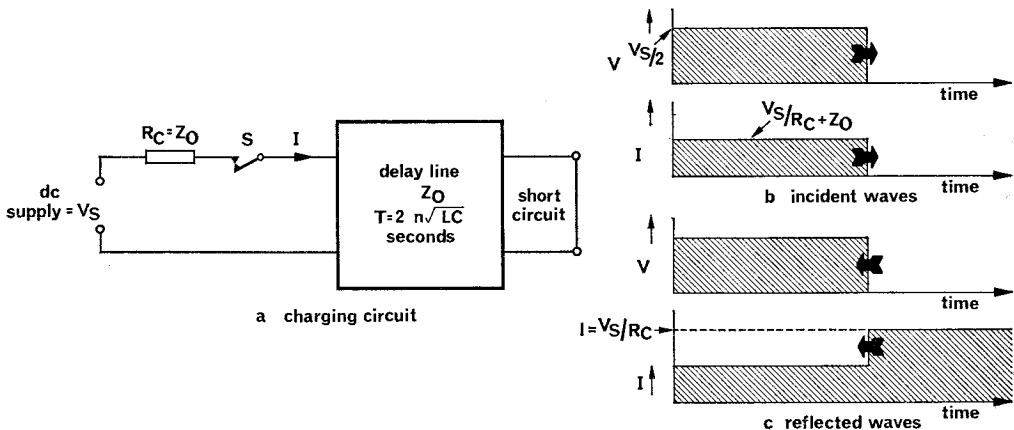


FIG 14. CHARGING A SHORT-CIRCUIT DELAY LINE

When switch S is closed, a voltage wavefront of amplitude $\frac{1}{2} V_s$ volts moves down the delay line towards the short-circuit termination. This voltage wavefront is accompanied by a current wavefront of amplitude $\frac{V_s}{R_C + Z_0}$ (Fig. 14b). When the wavefronts reach the short-circuited termination a reflected voltage wave of amplitude $\frac{1}{2} V_s$ volts and of *opposite polarity* to the incident voltage wavefront is reflected back towards the supply, reducing the voltage across the delay line to zero as it moves back. A current wavefront of the *same polarity* as the incident current wavefront is reflected from the termination causing a current of V_s/R_C amps to flow in the delay line. After $2n\sqrt{LC}$ seconds the wavefronts reach the supply terminals, the delay line voltage is zero at all points, and the current flowing is V_s/R_C amps. No further reflection occurs.

To form a pulse of duration $2n\sqrt{LC}$ seconds, the delay line is discharged through a load resistance R_L , equal to the Z_0 of the delay line. When the switch is connected as shown in Fig. 15a, the voltage across R_L falls instantly to $-\frac{1}{2} V_s$ volts. This voltage wavefront moves towards the short-circuit termination charging the delay line capacitors to $-\frac{1}{2} V_s$ volts. It is reflected from the termination with a change of polarity and returns to the input terminals after $2n\sqrt{LC}$

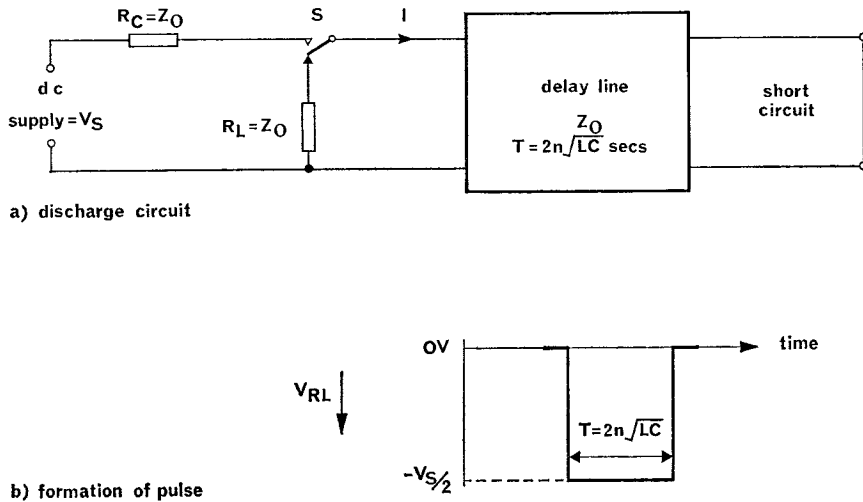


FIG 15. DISCHARGING A SHORT-CIRCUIT DELAY LINE

seconds, reducing the voltage across R_L to zero (Fig. 15b) as the delay line is completely discharged. Thus the energy stored in the magnetic field of the delay line is completely transferred to the load resistor forming a pulse of duration $2n\sqrt{LC}$ seconds and amplitude $-V_s/2$ volts.

Fig. 16a shows a circuit which can be used in the driver stage of a modulator to produce a narrow pulse from a wider input pulse. The output pulse duration is determined by the delay line constants.

A short-circuited delay line is connected in the anode circuit of a valve across a resistor R_L equal to the characteristic impedance of the line. The leading edge of the input pulse lifts the grid voltage above cut-off (t_1 in Fig. 16b), the valve conducts and the anode voltage falls immediately. This voltage change now moves down the delay line, is reflected, *with a change of polarity*, from the short-circuit termination and appears $2n\sqrt{LC}$ seconds later across R_L (t_2), forming a negative-going pulse.

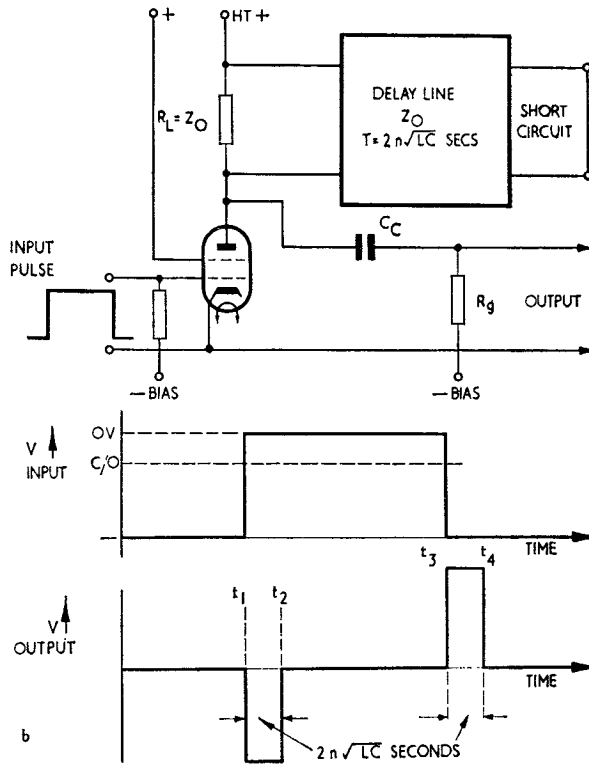


FIG 16. PULSE DRIVEN MODULATOR DRIVER STAGE

When the trailing edge of the input pulse cuts the valve off (t_3) the anode voltage rises and a positive-going pulse of duration $2n\sqrt{LC}$ seconds is formed. Thus two anti-phase voltage pulses can be taken as outputs via C_C R_g and, after limiting, the positive pulse can be applied to the power amplifier stage of the modulator.

Transmitter Power Supplies

An important part of the radar transmitter is the power supply unit which provides all the necessary voltages for the transmitter. The type of power unit will depend upon the requirements of the transmitter and the available inputs.

Because of the high e.h.t. voltages associated with the transmitter, precautions are taken to avoid damage to components due to excessive voltages and currents. A system of *interlocks* is always provided and delay circuits prevent e.h.t. voltages being applied before the valve heaters have warmed up. Switches are provided to break e.h.t. circuits if protective covers are removed. Blower motors and cooling fans are used to prevent rectifier and transmitter valves over-heating. Great care should be taken to avoid electric shock when working on a radar transmitter.

Pulse Compression Techniques

We know that for good range resolution a short duration pulse is required. However, to reproduce a short pulse accurately, the receiver bandwidth should be wide and this increases the receiver noise, making the echo less easily detectable and reducing the radar's range. This can be a disadvantage when the peak power of a transmitter is limited. With a long pulse, a narrow receiver bandwidth can be used and a greater range obtained; but a long pulse has poor range resolution.

To obtain the advantages of both a long and short pulse a technique called *pulse compression* has been developed. A long transmitter pulse is frequency modulated and the receiver is designed to compress this pulse, after i.f. amplification, into a much shorter one.

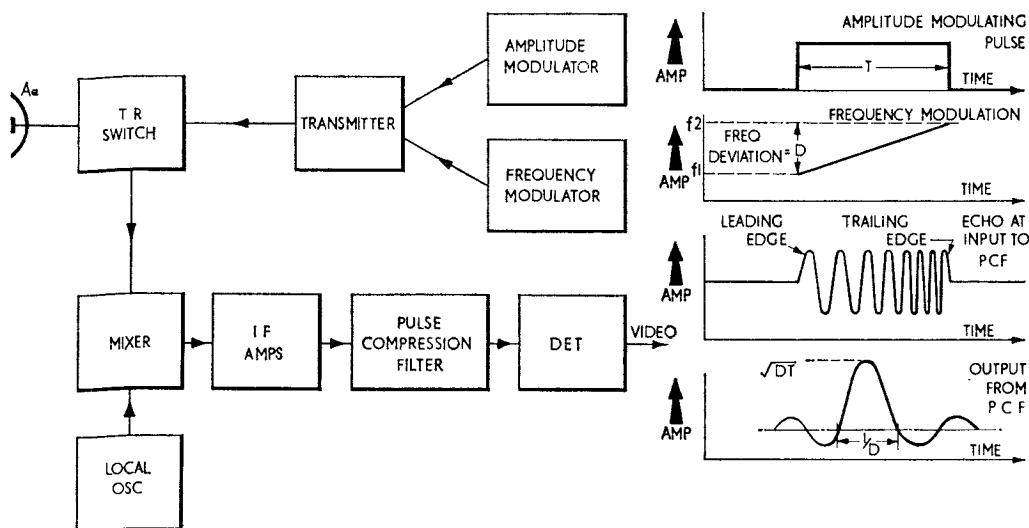


FIG 17. PULSE COMPRESSION

One method of pulse compression is illustrated in Fig. 17. The transmitter radiates a long, frequency modulated r.f. pulse of duration T and frequency deviation D . The echo from this pulse is received and amplified in the comparatively narrow-band i.f. amplifiers and is then passed through a *pulse compression filter* (p.c.f.). The velocity of propagation through the p.c.f. is proportional to frequency. Thus the higher frequencies at the trailing edge of the pulse are speeded up relative to the lower frequencies at the leading edge. In this way energy in the original long pulse is compressed into a shorter pulse.

The p.c.f. is a high pass filter and at centimetric wavelengths this could be a hybrid junction, rat-race or short slot with irises.

The main advantages of pulse compression are that a greater range can be obtained for a given pulse power and range measurements are more accurate than with simple amplitude modulation. Steep leading and trailing edges in the transmitted pulse are not essential.

CHAPTER 2

MODULATORS

Introduction

The modulator in a pulsed radar transmitter switches the supplies to the transmitter oscillator circuit on and off. This causes the transmitter oscillator to produce pulses of high-power r.f. for application to the aerial. Modulators can take various forms, depending on the frequency, power output, p.r.f. and pulse duration of the radar, but basically they all work on the principle of a valve being used as an on-off switch.

V.h.f. and u.h.f. radars can be modulated by positive pulses applied to the *grids* of the transmitter valves thus causing them to oscillate for the duration of the pulse. Another method is to place a modulator valve *in series* with the transmitter valves. By switching this modulator valve on and off at the grid, the e.h.t. to the transmitter valves is switched causing them to oscillate. In both grid and series valve modulators the pulse can be obtained from the master timing circuit which therefore controls the radar p.r.f. In low-power transmitters the master timing pulse may be applied directly to the grids of the transmitter valves or to the series modulator valve grid. In high-power transmitters the master triggering pulse may be applied to a *driver* or *sub-modulator* stage and then, for further amplification, to the modulator stage to ensure sufficient amplitude to operate the transmitter.

In most centimetric radar transmitters a high-power oscillator valve such as a magnetron is switched on and off by a negative e.h.t. pulse applied to its cathode, the anode block being earthed. The narrow, steep-sided e.h.t. pulse is often formed by a delay line or *pulse forming network*. The pulse forming network is charged from an e.h.t. power supply during the comparatively long period between pulses and then discharged rapidly through a gas-filled *switching valve* such as a thyratron or trigatron. The e.h.t. pulse is applied through a special *pulse transformer* to the magnetron. The switching valve is fired by a pulse from the master timer, which therefore controls the radar p.r.f. The pulse duration, however, is governed by the pulse forming network.

In this chapter we shall consider the principles of some of these types of modulator.

The Squegging Oscillator

The simplest form of pulsed radar transmitter is an oscillator which provides its own modulation, i.e. an oscillator which cuts itself on and off at a regular rate. Any type of oscillator will do this if the grid biasing CR circuit has a sufficiently long time constant.

Fig. 1 shows a conventional Hartley oscillator with automatic grid leak bias. The values of C and R are made large so that when oscillations start, C quickly acquires a large negative charge

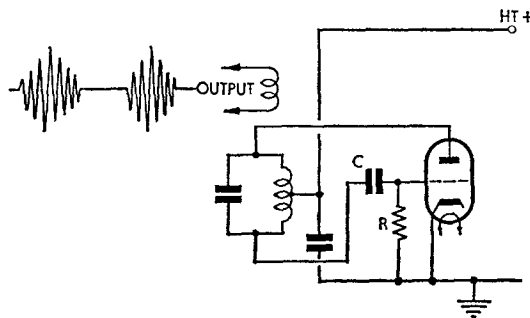


FIG. 1. THE SQUEGGING OSCILLATOR

due to grid current. This biases the valve beyond cut-off and oscillations cease. The charge on C then leaks slowly away through R and the grid voltage rises towards cut-off. When this point is reached another burst of oscillation occurs and the cycle repeats. The frequency of oscillation is determined by the tuned circuit and the p.r.f. by the value of grid CR.

Low-power Anode Modulation

Some low-power v.h.f. and u.h.f. radars use a single-stage modulator. This produces a positive pulse which is applied between anode and cathode of a triode oscillator to modulate directly the transmitter. The modulator circuit is often a blocking oscillator, which can produce e.h.t. pulses of sufficient amplitude using small valves and power supplies. The pulse duration is fixed by the type of transformer used but the p.r.f. may be varied by switching in grid leak resistors of different values.

A suitable free-running blocking oscillator circuit with waveforms is shown in Fig. 2. As well as being the transmitter modulator the blocking oscillator acts as the radar master timer

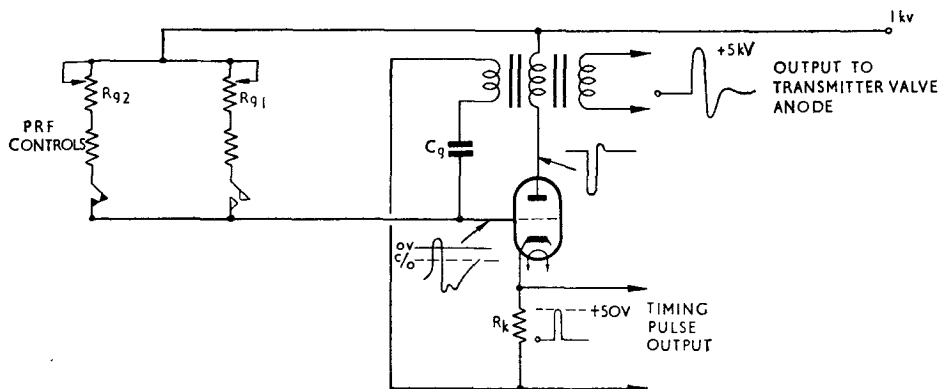


FIG. 2. SINGLE-STAGE BLOCKING OSCILLATOR ANODE MODULATOR

and the positive pulse developed across R_k can be used as a timing pulse for other circuits in the radar. The negative swing on the trailing edge of the e.h.t. output pulse due to transformer 'ringing' ensures that the transmitter is cut off completely at the end of the pulse.

Another circuit which can be used in this role is the critically-damped ringing oscillator (p. 148). The duration of the e.h.t. pulse produced by this circuit can be varied by adjusting the tuning slug of the anode inductor. As the ringing oscillator requires a pulse input it cannot be regarded as the radar master timer.

We shall see later that the blocking oscillator and ringing oscillator are widely used in more powerful modulators, either as a driver stage, or to trigger the discharge valve in a delay line modulator.

Hard-valve Series Modulator

This type of modulator is suitable for use in low-power v.h.f. and u.h.f. radars. A basic circuit is shown in Fig. 3a, and as shown in the equivalent circuit of Fig. 3b, the modulator valve V_1 acts as a low resistance switch in series with the oscillator and e.h.t. supply.

Initially the modulator valve is cut off by grid bias and this isolates the oscillator from the e.h.t. supply (switch open). When a trigger pulse of sufficient amplitude to overcome this bias is applied to V_1 grid the modulator valve conducts (switch closed). If the resistance of the conducting modulator valve is low compared with that of the oscillator circuit, most of the e.h.t. is developed across the oscillator and it oscillates for the period the modulator bias is lifted. Thus the pulse duration and p.r.f. of the radar are controlled by the trigger input pulse.

In some radars the grid bias CR of the oscillator valve is sufficiently long to cause the valve to cut off before the end of the trigger pulse (squegging action). In this case the p.r.f. is controlled by the input trigger but the pulse duration is governed by the oscillator grid CR.

In the series hard-valve modulator the modulator valve must be capable of passing a heavy current for the duration of the pulse; it must be able to withstand very high voltages and should have small inter-electrode capacitances to preserve pulse shape. Two such valves in parallel will double the current that can be passed.

Grid Modulation

Another type of modulation circuit using a cathode follower as a modulator valve is shown in Fig. 4. The grid of the oscillator valve V_2 is connected to the cathode of the modulator valve V_1 and between input pulses V_1 passes a small current. The voltage at V_1 cathode is therefore

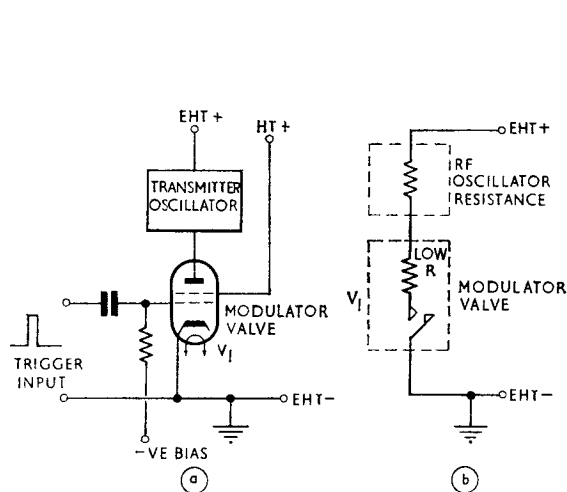


FIG. 3. BASIC SERIES HARD-VALVE MODULATOR

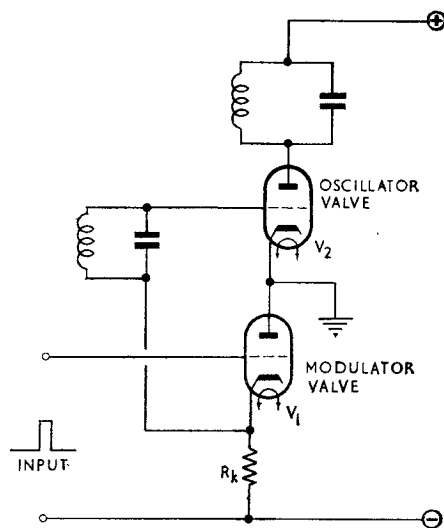


FIG. 4. BASIC CATHODE FOLLOWER GRID MODULATOR

sufficiently negative to hold V_2 cut off. When a positive pulse is applied to V_1 grid, V_1 conducts heavily and its cathode voltage rises sharply, cutting V_2 on. V_2 oscillates until the trailing edge of the input pulse causes it to cut off.

Driver-Power Amplifier Modulator

In the driver-power amplifier modulator a pulse is produced by the master timer at a low power level. It is then amplified and shaped by a driver stage (sometimes called a sub-modulator stage) and applied to a power amplifier before being finally used to modulate the transmitter (Fig. 5).

The master timer. This can be one of the pulse circuits already described, e.g. a cathode-coupled multivibrator, ringing oscillator, or blocking oscillator. If a very stable p.r.f. is required

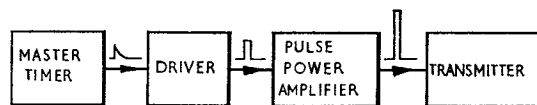


FIG. 5. BLOCK DIAGRAM OF A DRIVER-POWER AMPLIFIER MODULATOR

an RC oscillator or crystal controlled oscillator followed by pulse shaping circuits may be used. This circuit controls the p.r.f. and for a radar with a low p.r.f. a circuit with a suitable mark-to-space ratio is required. It need not provide much power.

The driver stage. This must provide a large negative voltage between pulses to keep the power amplifier stage cut off, a large positive voltage during the pulse and a fairly large peak power (several hundred watts). The pulse output of the driver must be of accurate duration, steep-sided and of constant amplitude. Such a pulse can be produced by using a pulse forming network with the driver stage.

A driver stage using a blocking oscillator and an open-circuited pulse forming network or delay line is shown in Fig. 6a. A normal blocking oscillator is used but the pulse duration is controlled by the pulse forming network and not by the normal capacitance and inductance of

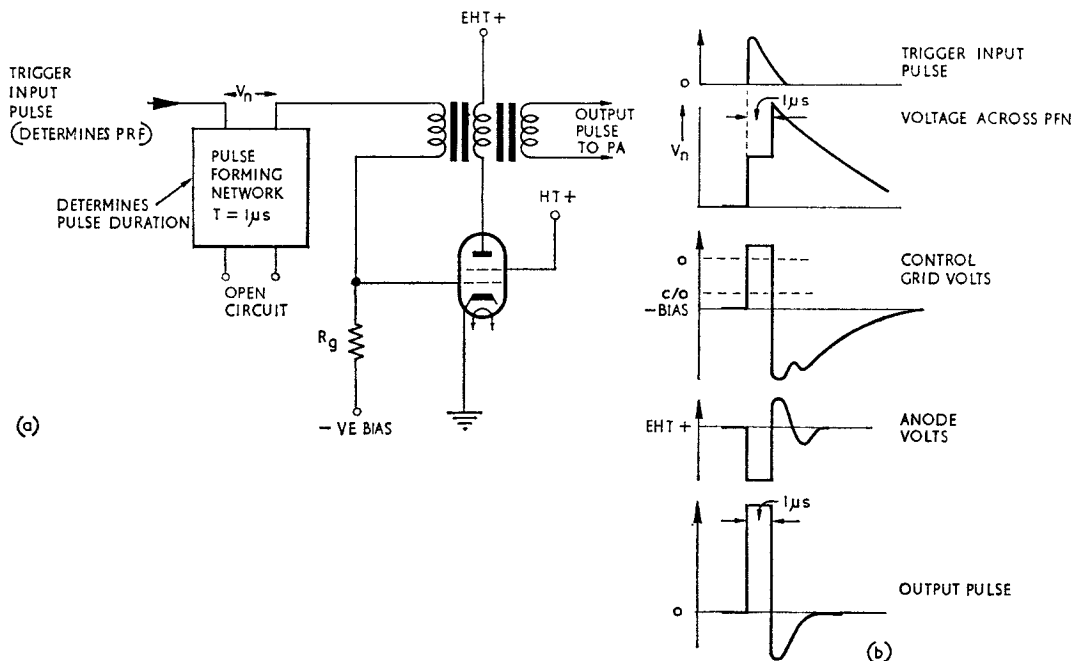


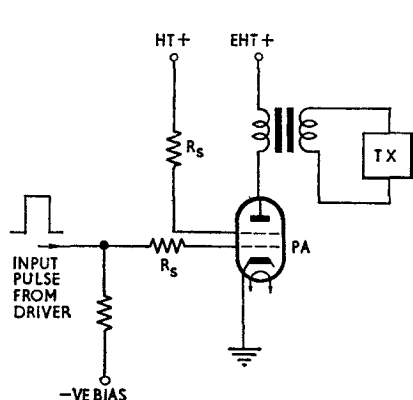
FIG. 6. BLOCKING OSCILLATOR DRIVER WITH PULSE FORMING NETWORK

the transformer. Positive input trigger pulses are applied to the blocking oscillator grid and positive pulses of $1\mu s$ duration, controlled by the pulse forming network and p.r.f. controlled by the master timer are taken to the power amplifier stage.

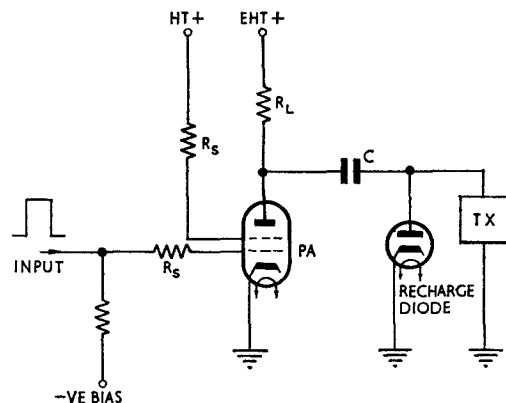
Initially the valve is cut off by the negative bias on its grid. The positive trigger pulse is coupled through the pulse forming network and transformer to the grid which is raised above cut-off and conducts heavily. The anode voltage falls and the grid voltage rises sharply due to blocking oscillator action. Grid current flows through the pulse forming network and causes a sudden increase in voltage across its terminals. Thus a wavefront moves towards the open-circuit end, is reflected in phase and causes a second increase to appear at the input end of the pulse forming network $1\mu s$ later. This starts a second switching action that ends the pulse. The pulse forming network then discharges through R_g .

Power amplifier stage. This consists of a large power valve which is biased beyond cut-off. Sometimes two or more tetrodes connected in parallel are used. The pulse from the driver stage lifts the grid well into grid current causing a very large anode current pulse.

The p.a. stage may be transformer-coupled to match its output impedance to the transmitter impedance, the large secondary voltage pulse being used as the e.h.t. supply to the transmitter (Fig. 7a). Grid and screen stopper resistors (R_s) are often necessary to prevent parasitic oscillations especially when valves are used in parallel.



(a) TRANSFORMER COUPLING



(b) RC COUPLING

FIG. 7. PULSE POWER AMPLIFIERS

Alternatively, the p.a. stage may be coupled to the transmitter by RC coupling (Fig. 7b). Between pulses the coupling capacitor C is charged to e.h.t. via R_L and the recharge diode. When the p.a. stage conducts heavily its anode voltage falls and C discharges via the transmitter and p.a. valves. Energy stored in C is thus transferred to the transmitter. In some modulators the discharge current of C flows through the primary of a step-up transformer to produce a higher e.h.t. for the transmitter. In this case the diode is used to damp out the transformer 'ring'.

Sometimes a gas discharge valve is used as the p.a. stage. Because of its low conducting resistance, losses are less and smaller triggering power is required. In this case the pulse duration is determined by the discharge time of C and not by the input trigger pulse.

High-power Delay Line Modulators

Most high-power radar transmitters using a magnetron are modulated by an e.h.t. pulse formed by a delay line pulse forming network. The pulse forming network is charged through an impedance during the interval between pulses and in response to a trigger pulse the delay line is discharged rapidly through a special triggering switch and the transmitter. It is important that the characteristic impedance of the pulse forming network is matched to the transmitter impedance and this is normally achieved by a special pulse transformer. A block diagram of the system is shown in Fig. 8. The p.r.f. of the radar is governed by the triggering pulses, which are usually produced by a

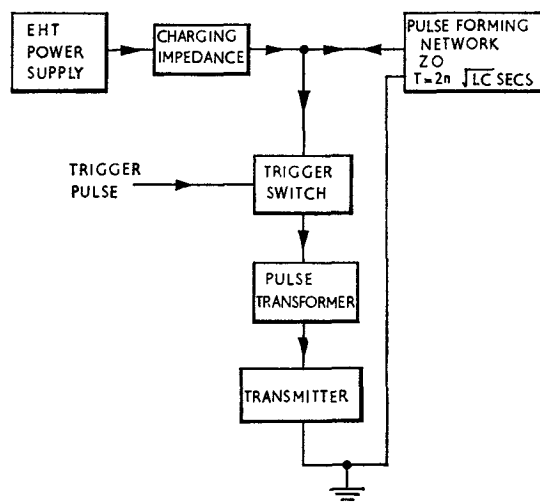


FIG. 8. OUTLINE OF A HIGH-POWER DELAY LINE MODULATOR

blocking oscillator or ringing oscillator, but the pulse duration is determined by the constants of the pulse forming network.

The current drawn from the e.h.t. power supply and stored in the pulse forming network during the relatively long inter-pulse period, is small compared with that supplied by the pulse forming network during the short pulse period. For example, using the values given in Fig. 9, the charging current is 50 mA and the discharging current 50A.

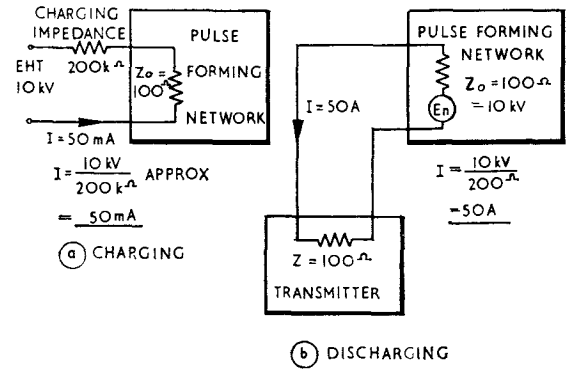


FIG. 9. CHARGING AND DISCHARGING CURRENTS

Charging Through a Resistor

The outline of a modulator using an open-circuit delay line as a pulse forming network, charged through a resistance is shown in Fig. 10a. Initially the delay line is uncharged and the discharge switch is also open. When e.h.t. is applied, the delay line charges to the value of the e.h.t. supply via the charging resistor R_{CH} . As R_{CH} is very large compared to the Z_o of the delay line, the individual steps due to mis-match are negligible and are not shown in Fig. 10b. The circuit behaves as a simple CR and V_n rises exponentially (A to B).

When the switch is closed at instant B, then provided the transmitter impedance equals the characteristic impedance of the delay line, the voltage across the transmitter rises to $\frac{\text{e.h.t.}}{2}$ and V_n

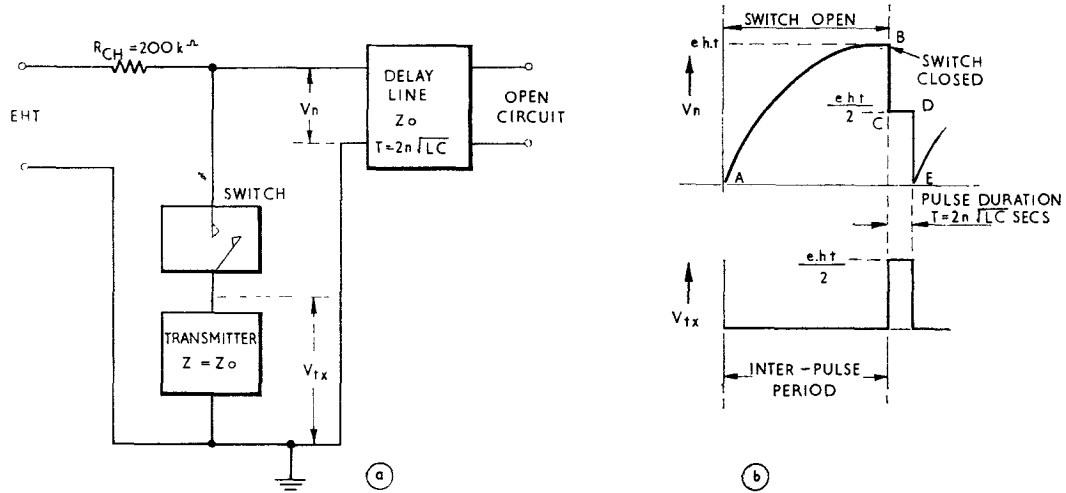


FIG. 10. RESISTANCE CHARGED DELAY LINE

falls to $\frac{\text{e.h.t.}}{2}$ (BC). This fall is reflected in-phase from the open circuit termination, and the transmitter voltage and V_n fall to zero after $2n\sqrt{LC}$ seconds (C to E), where n is the number of sections and L and C are the inductance and capacitance per section of the delay line. The switch is now opened, the delay line recharges and the cycle repeats.

Thus a narrow, well-shaped pulse of duration $2n\sqrt{LC}$ seconds and amplitude $\frac{\text{e.h.t.}}{2}$ volts is applied to the transmitter.

Triggering Valves

In practice the switch shown in Fig. 10a is usually a gas-filled valve such as an *ignitron*, a *thyatron* or a *trigatron*. An external trigger pulse is applied to the control grid and causes the valve to conduct (switch closed); when the delay line is discharged the gas-filled valve de-ionises (switch open) allowing the delay line to recharge.

The triggering device must have a low conducting resistance and be capable of passing a heavy current and withstanding high voltage. Once triggered it must continue to conduct until the delay line is discharged, then rapidly return to an open-circuit. It should be reasonably robust with a long life and constant performance under varying temperature and pressure conditions.

The types of pulse forming network triggering devices commonly used in radar have been discussed on p. 160. The *ignitron* is capable of passing very heavy currents and is generally used in long-range ground radars. The *trigatron* is a cold-cathode valve filled with a mixture of argon and oxygen to provide a low conducting resistance and short de-ionising time. It is normally used in airborne radars. The hydrogen-filled *thyatron* is the most widely used triggering device. It can pass heavy currents and is suitable for the very rapid switching action required with narrow-pulse high-p.r.f. radars.

Pulse Transformer

The transformer between the pulse-forming network and the transmitter in Fig. 8 is necessary for two reasons. Firstly, it matches the transmitter magnetron impedance (approximately 1 k Ω) to the characteristic impedance (Z_0) of the pulse-forming network (typically 100 Ω). In some radars the modulator unit is placed at a distance from the transmitter unit and a *pulse cable* is required to carry the e.h.t. modulating pulse to the transmitter. The pulse transformer then matches the Z_0 of the pulse-forming network to that of the cable.

Secondly, the pulse transformer steps up the pulse-forming network voltage applied to the magnetron. This allows the pulse-forming network to be operated at a much lower voltage than the magnetron thus reducing insulation problems in most of the modulator circuit.

The pulse transformer must faithfully reproduce the pulse formed by the pulse-forming network. A high-permeability core is used and eddy current losses are kept to a minimum. The primary and secondary windings are coupled closely together and the inter-turns capacitance is small. The pulse transformer must be capable of handling high current and voltage without core saturation or insulation breakdown. It is often mounted in an oil-filled container.

The magnetron cathode is usually connected to one side of the heater and to avoid the need for a heavily insulated heater transformer the pulse transformer has two identical secondary windings. These *bifilar* secondary windings have the same e.h.t. voltage induced in each. The low-voltage magnetron heater supply is applied via these windings as shown in Fig. 11 and as the two lower ends of the bifilar windings are at the same voltage a normally-insulated heater transformer can be used. Balancing capacitors C_1 and C_2 are sometimes used to ensure equal voltages in the two secondary windings.

When the magnetron is oscillating, the cathode is bombarded by electrons which provide

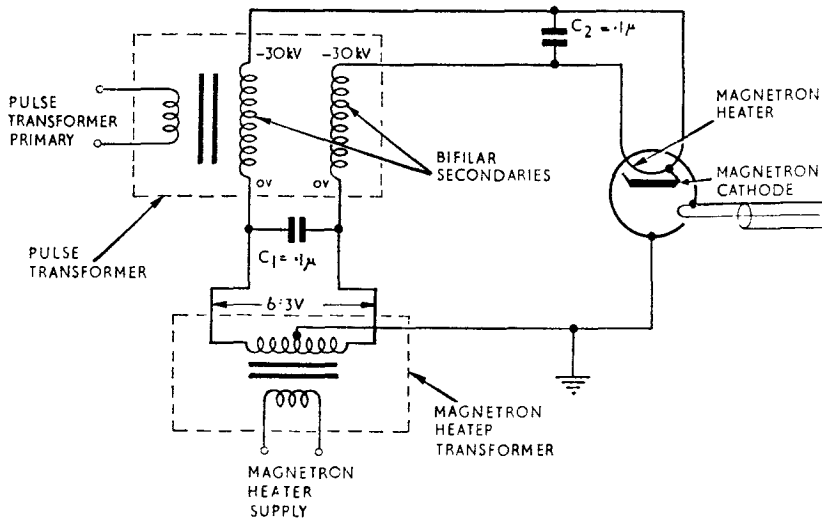


FIG. 11. MAGNETRON PULSE TRANSFORMER WITH BIFILAR SECONDARIES

heat. Thus, to prevent the cathode being over-heated, the heater voltage may be reduced by increasing the number of primary turns on the heater transformer once the magnetron is operating (Fig. 12). Sometimes an additional secondary winding on the pulse transformer is used to provide a synchronising pulse for other circuits in the radar and an anti-blocking pulse for the radar receiver.

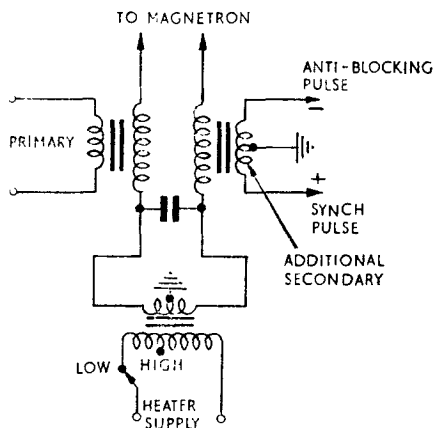


FIG. 12. OUTLINE OF A PRACTICAL PULSE TRANSFORMER

Charging Through a Choke

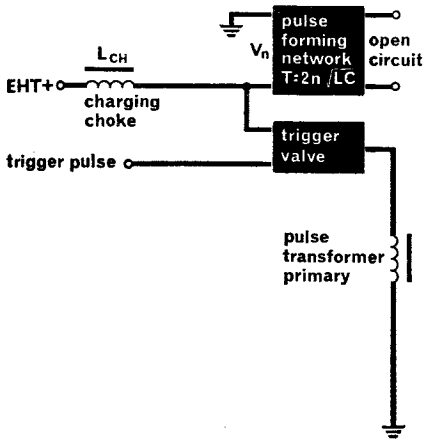
When a pulse-forming network is charged through a resistance, the voltage developed across the pulse transformer primary winding is only half the e.h.t. supply voltage. Thus to produce a certain amplitude pulse an e.h.t. supply unit giving twice the pulse voltage is required.

This disadvantage is overcome by charging the pulse-forming network through a large value choke (L_{CH}) as in Fig. 13a. L_{CH} is in series with the capacitors in the pulse-forming network and, together, the choke and capacitors form a resonant series circuit. When

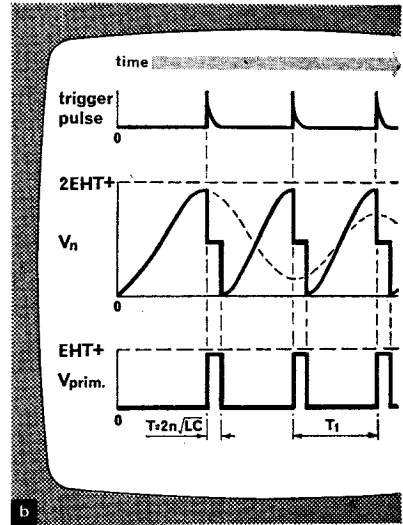
the e.h.t. supply voltage is applied, this circuit 'rings' and the pulse-forming network capacitors charge to almost *twice* the e.h.t. supply voltage (Fig. 13b). At the instant when V_n reaches this peak the trigger pulse is applied to the trigger valve, which ionises, and the pulse-forming network discharges, preventing the remainder of the ring (shown dotted in Fig. 13b).

When the line is discharged, half the pulse-forming network voltage of 2 e.h.t. is developed across the pulse transformer primary (V_{prim}) and half across the network itself. Thus an output pulse of amplitude equal to the e.h.t. supply voltage is obtained.

In the circuit of Fig. 13 the resonant frequency of the circuit formed by the charging choke and the pulse-forming network capacitors must be *half* the radar p.r.f. This ensures that the trigger pulse coincides with the peak of V_n . The triggering time can, in fact, be varied *slightly* around this peak but if a different p.r.f. is required, as in some multi-range radars, a hold-off



a

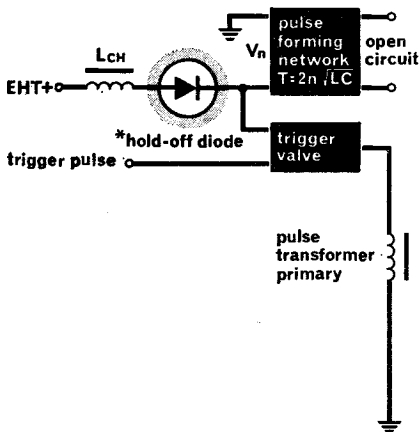


b

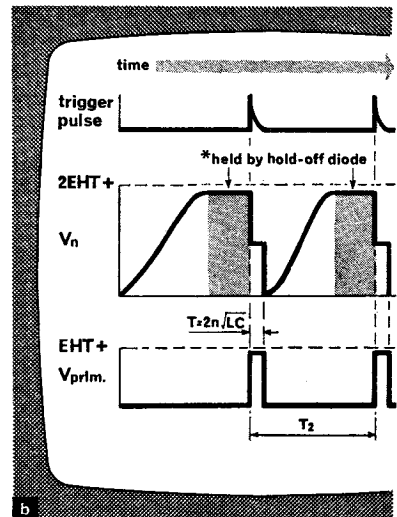
FIG 13. RESONANT CHOKE CHARGING

diode must be used (Fig. 14). This diode prevents the pulse-forming network discharging during the second half cycle of the 'ring' and 'holds' the voltage at its peak value until the line is triggered. The p.r.f. may now be varied; for example, T_2 in Fig. 14 is greater than T_1 in Fig. 13.

The pulse-forming network can be charged positively with respect to earth, as in Figs. 13 and 14, or negatively. The e.h.t. pulse applied to the magnetron cathode must, of course, be negative, the anode block being earthed.



a



b

FIG 14. RESONANT CHOKE CHARGING WITH HOLD-OFF DIODE

Overswing Diodes

By common usage, the term 'overswing diode' has been applied to two different circuit functions. To differentiate between the two, the terms 'overswing diode (line discharge)' and 'overswing diode (anti-ringing)' are used.

Overswing diode (line discharge). So far we have assumed a perfect match between the pulse-forming network and the primary of the pulse transformer, so that when the line is discharged the available voltage is split evenly between the two. In practice it is not always possible to get a perfect match and this introduces certain problems.

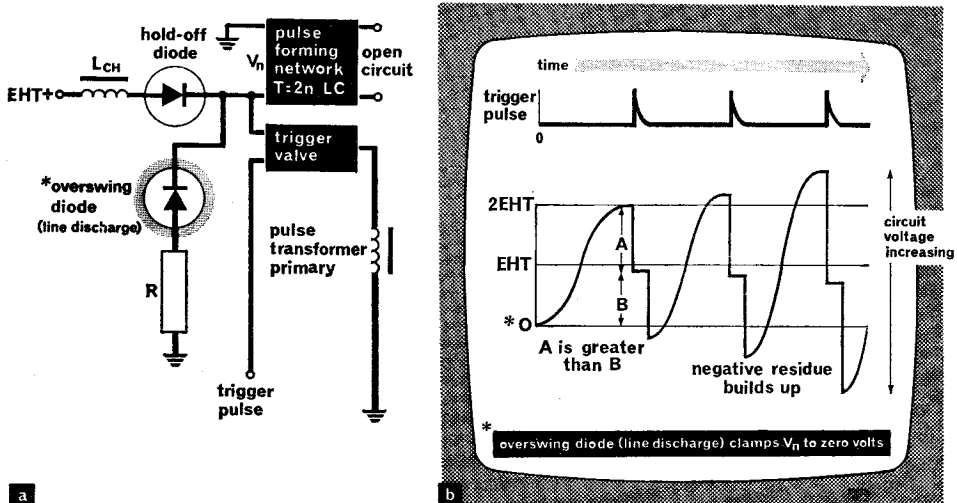


FIG 15. USE OF OVERSWING DIODE (LINE DISCHARGE)

If the impedance of the pulse transformer primary is *less* than that of the pulse-forming network, the distribution of voltage between the two is unequal, the greater part of the voltage being developed across the pulse-forming network. In a perfectly matched line the pulse-forming network is completely discharged at the end of the pulse, all the energy having been expended. For the mismatched circuit described above, the pulse-forming network 'overswings' on discharge, because its energy cannot be fully expended in the load, and it is then left with a *negative* charge (Fig. 15b).

This charge acts in series with the applied e.h.t. and may build up progressively on each cycle to such a high value that circuit components are damaged. An overswing diode (line discharge) is included in the circuit to prevent this happening. It is connected as shown in Fig. 15a so that it removes any negative over-swing charge from the pulse-forming network at the end of each pulse. The energy is expended in the resistor R .

If the mismatch is the other way round, i.e. if the impedance of the pulse transformer primary is *greater* than that of the pulse-forming network, the network will be left with a *positive* residue at the end of each pulse. The characteristics of the circuit are such that this charge is removed by the trigger valve before the next cycle commences.

Overswing diode (anti-ringing). This diode

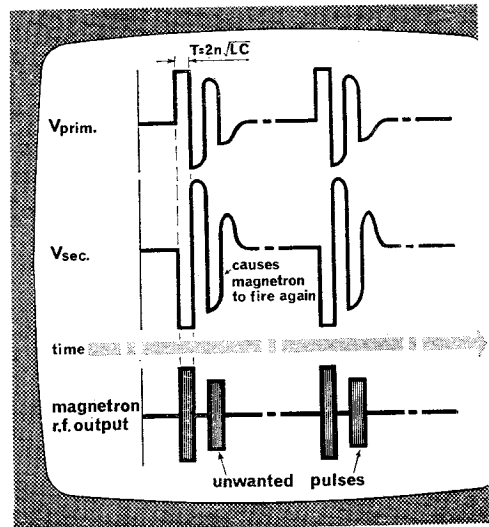


FIG 16. EFFECT OF RINGING IN PULSE TRANSFORMER

is included to reduce 'ringing' in the primary of the pulse transformer. Severe ringing would cause the magnetron to double pulse (Fig. 16) and this, in turn, would cause confusion of echoes on the radar display.

The overswing diode (anti-ringing) is connected across the primary of the pulse transformer as shown in Fig. 17a. It is connected such that it limits the negative swing of the primary waveform and ringing is greatly reduced. The magnetron now fires only on the wanted pulse (Fig. 17b). This anti-ringing circuit is common in equipments which use short duration pulses.

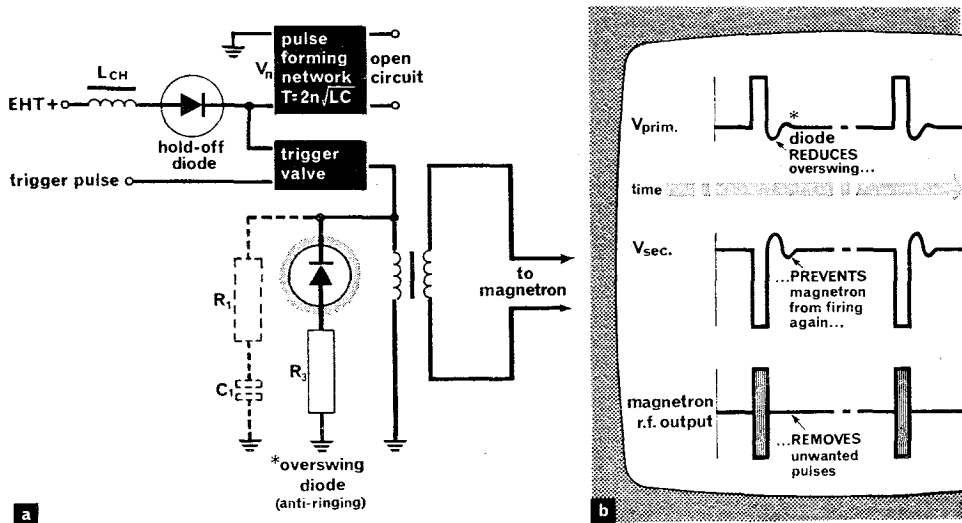


FIG 17. USE OF OVERSWING DIODE (ANTI-RINGING)

Fig. 17a also shows a network $R_1 C_1$ across the pulse transformer primary. The purpose of this network is to slow down the rate at which power builds up in the magnetron. A gradual build-up, over about 10 nanoseconds, allows the TR cell to ionise completely and reduces the energy content of the initial 'spike' which gets through to the receiver (see p. 309).

Other Types of Modulator

We have considered details of several types of modulator used in pulsed radar transmitters. The modulator is designed to suit the particular requirements of the radar and two radars will not necessarily have identical modulators.

Pulse-forming networks are widely used because they can produce a modulating pulse of the required shape, amplitude and duration. The pulse duration can be conveniently altered by switching in or out sections of the pulse-forming network. This is often done in multi-range radars where the pulse duration and p.r.f. are varied to suit each range. Because of the high voltage and current some pulse-forming networks are immersed in an oil bath.

We have seen that a pulse-forming network can be operated by an input pulse or by a d.c. e.h.t. supply. It can also be charged by an a.c. supply, through a diode, if the p.r.f. is synchronised to the frequency of the power supply.

A *saturable reactor* can be used as a switching device with a pulse-forming network to form a modulating pulse. It has the advantage of being reliable and produces a high amplitude pulse from a comparatively low e.h.t. supply.

Solid state modulators are used in low- and medium-power mobile radars where their compactness and light weight are important features.

Simplified Pulsed Radar Transmitter

Fig. 18 shows the simplified circuit of a centimetric radar transmitter incorporating some of the circuits and components discussed in this chapter. The blocking oscillator V1 produces the master timing pulses for the equipment at the required p.r.f. and these are used to trigger the display time-base and other timing circuits. The grid waveform of the blocking oscillator is used to trigger the ringing oscillator V2 which produces the high voltage trigger pulses needed to ionise the trigatron and isolates the blocking oscillator from the rest of the modulator. When the trigatron ionises, the pulse forming network (which has been charged via the charging choke and the hold-off diode) discharges and develops half its voltage across the pulse transformer primary. This voltage is stepped up in the bifilar secondary windings and applied to the magnetron cathode. The magnetron fires and produces the high power r.f. pulse in the waveguide output.

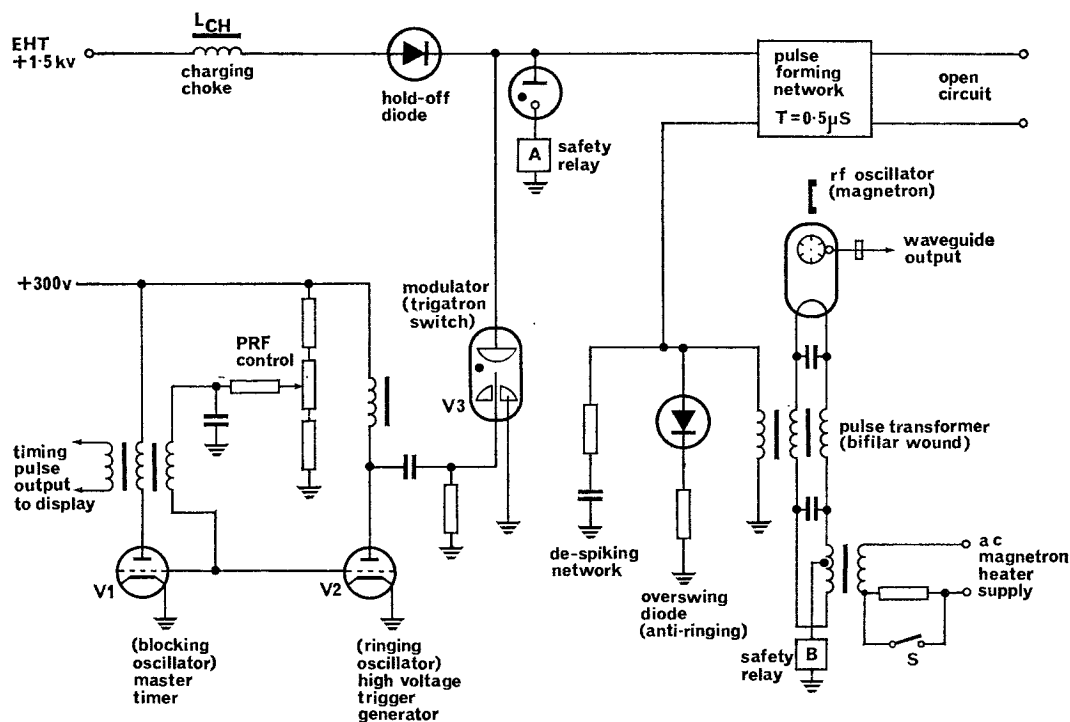


FIG 18. SIMPLIFIED PULSED RADAR TRANSMITTER

Several protective devices are included in the circuit. Safety relay A is operated if the EHT voltage stored on the line becomes excessively high and ionises the discharge tube. Safety relay B operates if the mean magnetron current exceeds the design value. Both of these relays switch off the EHT supply when they operate. The switch S in the magnetron heater circuit opens after a few minutes operation to reduce the magnetron heater current and prevent overheating of the cathode. The overswing diode (anti-ringing) prevents double pulsing of the magnetron and the de-spiking network slows down the build up of power to prevent the mixer crystal from being burnt out.

CHAPTER 3

TRANSMITTER POWER SUPPLIES

Introduction

The power contained in the r.f. pulse radiated by the radar transmitter comes originally from the power source supplying the radar. For ground radars this source may be single-phase a.c. mains (240V 50 c/s) or three-phase mains (415V 50 c/s); for airborne radars it can be low d.c. (28V), medium d.c. (112V) or three-phase a.c. (200V 400 c/s), depending upon the type of aircraft.

The transmitter *power unit* must convert this primary supply into the many voltages required to operate the transmitter. Where the primary source is d.c. an inverter or motor-generator is used to convert it to a suitable a.c. supply and transformers are then used to give the required input voltages to the rectifier circuits.

Typical transmitter voltage requirements are: low-voltage a.c. for valve heaters; positive h.t. of various values for valves; negative h.t. for biasing arrangements; and e.h.t. of several thousand volts for the modulator. In addition, a.c. or d.c. supplies are required to operate ancillary machinery such as blower motors and relays. The voltage requirements will depend upon the design of the transmitter, and the power units will therefore differ with each radar.

As well as converting the primary source into suitable value alternating and direct voltages the power unit must ensure that the voltages do not vary. Voltage regulators and stabilisers therefore form a part of most power units. These transmitter power unit requirements are summarised in Fig. 1.

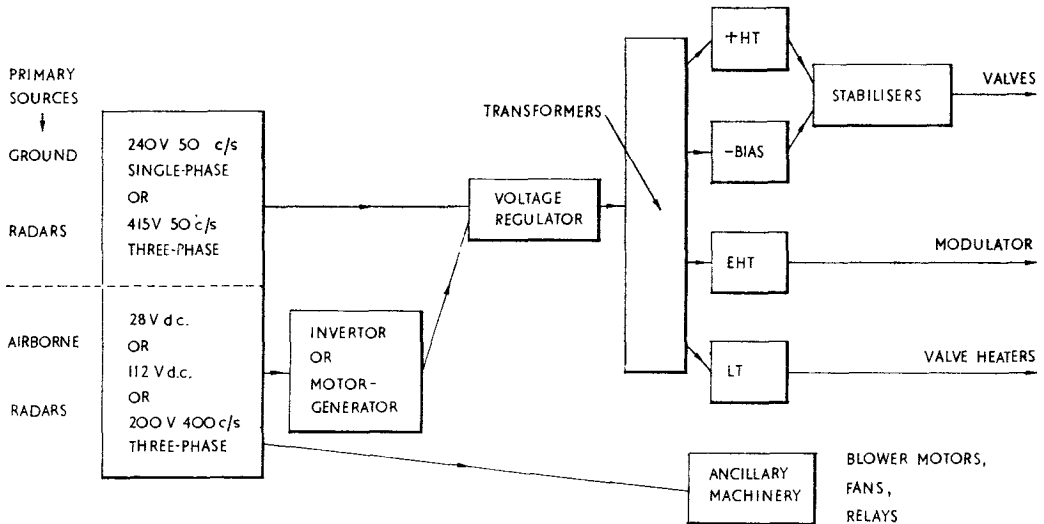


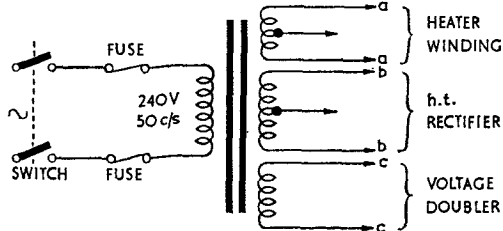
FIG. 1. OUTLINE OF A RADAR TRANSMITTER POWER UNIT

In this chapter we shall consider circuits used in radar power supply units to provide the necessary voltages and to stabilise and regulate these voltages. Component protection and interlock circuits will also be discussed.

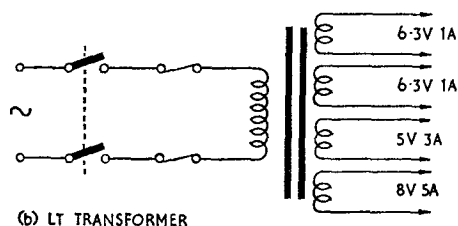
Power Unit Transformers

The power unit of a radar transmitter usually has several transformers to change the input a.c. voltage to the value required by rectifier and other circuits. Normally there are several

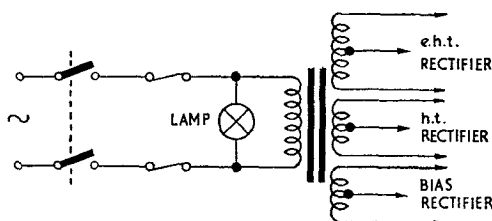
secondary windings on the transformer, each with a suitable turns ratio. Sometimes the transformer has an l.t. secondary winding to supply valve heaters and one or more h.t. windings feeding the rectifier (Fig. 2a). More frequently separate h.t. and l.t. transformers are used to provide the necessary power (Fig. 2b and c).



(a) COMBINED LT AND HT TRANSFORMER



(b) LT TRANSFORMER



(c) HT TRANSFORMER

FIG. 2. TYPES OF SINGLE-PHASE TRANSFORMER

When the primary source is three-phase a.c., separate transformers can be used for each phase or a single three-phase transformer may be employed. In the latter case the three phases may be connected in delta, when the phase voltage equals the line voltage (Fig. 3a), or in star (Fig. 3b) when the line voltage is $\sqrt{3}$ times greater than the phase voltage. In both cases the power output is the same and equals $\sqrt{3} V_L I_L$ watts, assuming unity power factor. Often the primary windings are connected in delta and the secondaries in star (Fig. 3c).

In some radars a two-phase supply is required to operate certain components (e.g. a two-phase induction motor). This may be obtained from a three-phase supply by using a *Scott connected* transformer. Two transformers, one with a centre-tapped secondary, are connected as in Fig. 4. When a three-phase input is applied to terminals ABC a two-phase supply is available at the output terminals.

Gas-filled Rectifier Valves

Vacuum and semiconductor diodes are used in low and medium power rectifier circuits but where a high current is required from the rectifier, gas-filled valves are often employed. Mercury-vapour valves have a high current-capacity, low conducting resistance and good regulation. The mercury vapour inside the valve envelope is obtained by a small bead of mercury vaporising as the valve heats up. The valves must be carefully handled to avoid splashing the mercury over the electrodes. The heater must always be allowed to warm up to the correct temperature before the h.t. supply is switched to the anode. This enables the mercury to vaporise and prevents high-velocity ions damaging the cathode. To enable this to be done automatically a delay circuit is always incorporated in a gas-filled valve rectifier circuit.

Semiconductor devices known as *thyristors* (sometimes called silicon controlled rectifiers) are capable of passing heavy currents, and are replacing gas-filled and mercury vapour valves and even mercury-arc rectifiers and ignitrons.

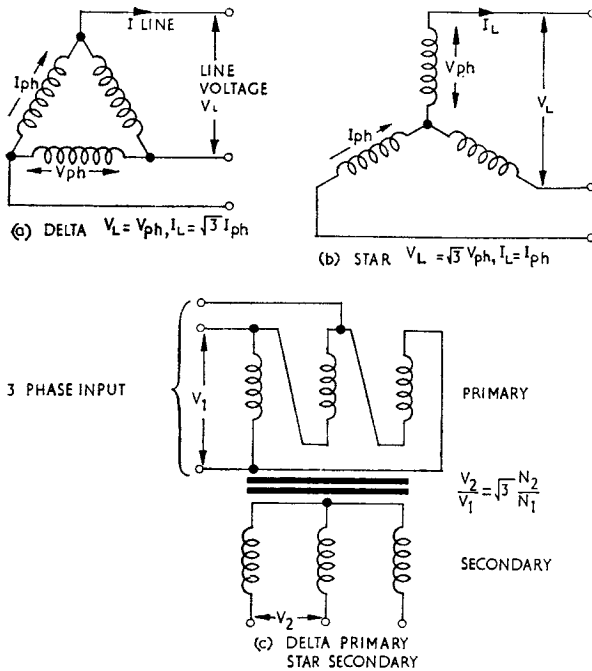


FIG. 3. THREE-PHASE CONNECTIONS

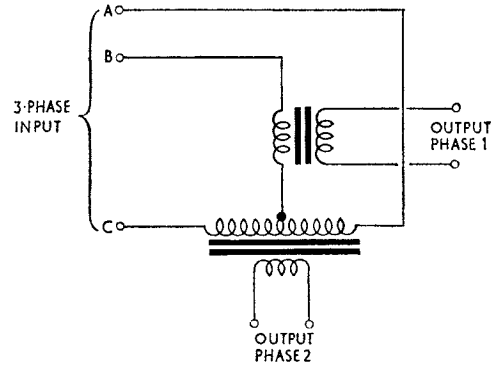


FIG. 4. SCOTT CONNECTED TRANSFORMER

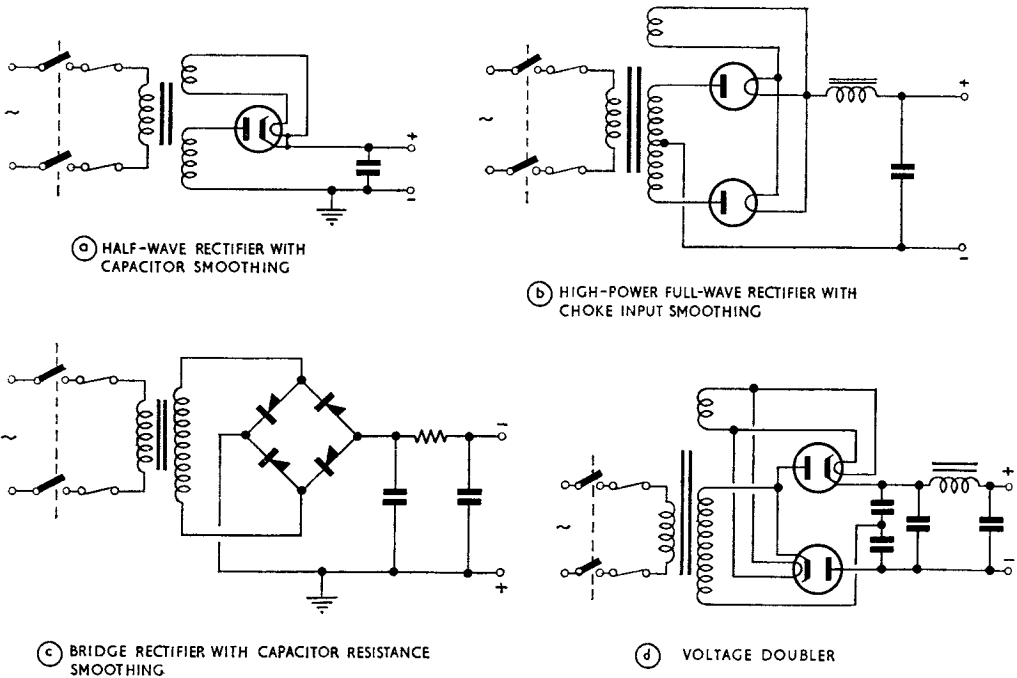


FIG. 5. SINGLE-PHASE RECTIFIER CIRCUITS

Rectifier Circuits

The main types of single-phase rectifier circuits have been considered in Part 1. Fig. 5 shows some rectifier and smoothing circuits commonly used in radar transmitter power units. When the current drawn from the rectifier is small, RC filters are used in preference to the heavier LC filter.

Ground and airborne radars often employ a three-phase primary source. This has the advantage over a single-phase supply of greater efficiency and, since the ripple frequency is higher, smaller and lighter smoothing components can be used. A simple three-phase half-wave rectifier circuit and waveforms are shown in Fig. 6. Notice that the output ripple frequency is three times the supply frequency.

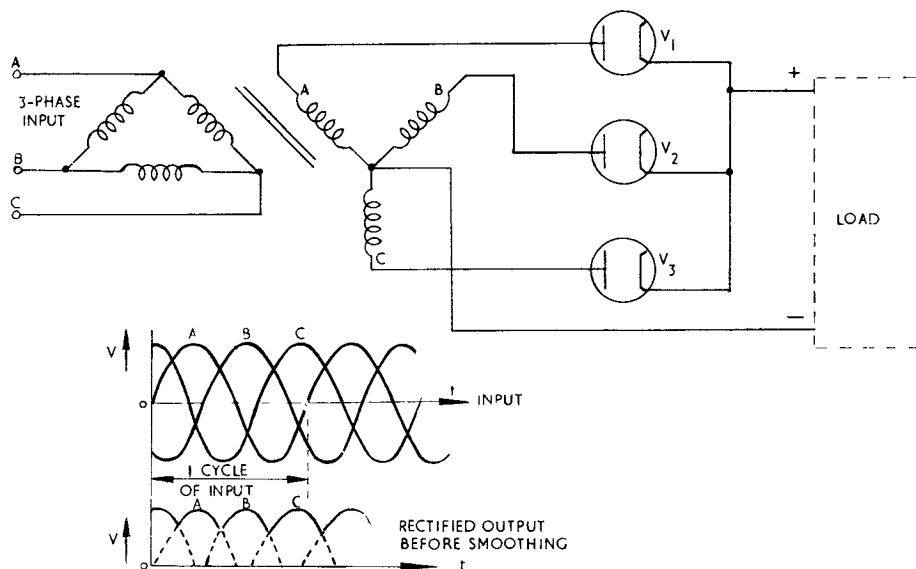


FIG. 6. THREE-PHASE HALF-WAVE RECTIFICATION

Fig. 7 shows a three-phase full-wave rectifier circuit and waveforms. From the input waveforms it can be seen that for a time while phase A is positive phases B and C are negative. During this time current flows through V_1 to the positive side of the load, through the load and back through V_5 and V_6 to points B and C. Notice that the ripple frequency on the output d.c. is six times the input frequency.

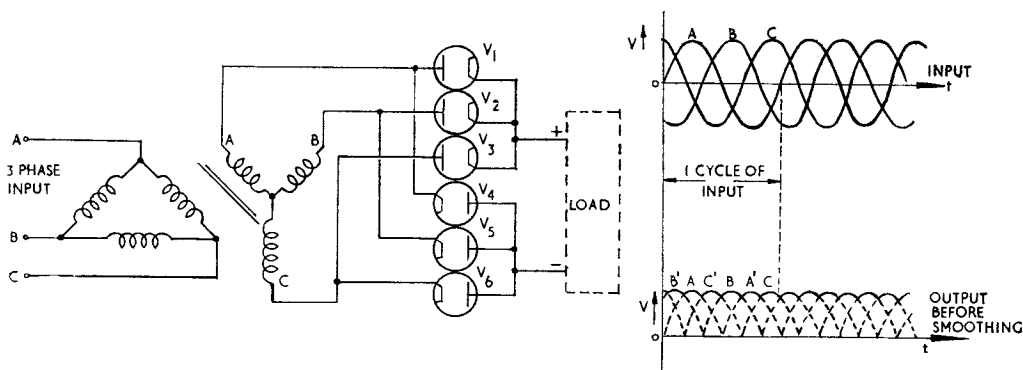


FIG. 7. THREE-PHASE FULL-WAVE RECTIFICATION

Mercury-arc rectifier. Some long-range high-power ground radars require more power than can be provided by normal hard or soft valve rectifier circuits. A *mercury-arc rectifier* with a three-phase input can supply currents of up to 1,000 amps.

A glass envelope contains several *carbon anodes* and a pool of mercury which forms the cathode (Fig. 8). A *striker anode* is mounted close to the mercury pool and is supplied with a voltage. The striker anode is spring-loaded and when an electromagnet is energised it pulls the striker into the mercury pool causing a current of about 5 amps to flow. The electromagnet is now short-circuited and the striker springs out of the mercury, forming an arc. This arc causes some of the mercury to vaporise and a current flows between the cathode and the main anodes connected to the secondaries of a three-phase transformer.

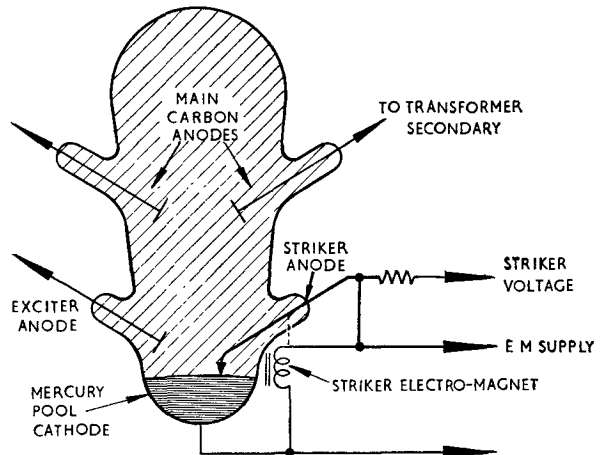


FIG. 8. MERCURY-ARC RECTIFIER

Positive ions produced by the ionised mercury strike the cathode and maintain ionisation. A white-hot spot forms on the surface of the mercury and a low-resistance path is formed between this spot and the conducting anodes, enabling a heavy current to flow. As the three-phase a.c. input makes each anode in turn positive, so the current path from the cathode switches from one anode to the next. Once the main arc has been struck the voltage on the striker anode is removed.

If the load current were to fall below a certain level the arc would be extinguished and the ignition process would have to take place again. To avoid this an *exciter anode*, supplied from a separate transformer, maintains the cathode spot in spite of changes in the d.c. load current.

Thermal Delay Switch

A thermal delay switch is a device used to delay the application of a voltage to a circuit for a certain time after switching on. It is used, for example, in mercury vapour rectifier circuits and in circuits where an e.h.t. voltage would strip the cathode if the cathode were not freely emitting electrons.

The basic construction of a thermal delay switch is shown in Fig. 9. Two small lengths of *different* metals are joined together and held firmly at one end. When this *bimetallic strip* is heated by current flowing through a heater element one metal strip expands more than the other and the strip warps as shown. After a specified time the strip is sufficiently warped to close a set of switch contacts which makes the required circuit.

Thus a time delay is imposed between applying current to the valve heaters and the application of h.t. or e.h.t. to the valve anode.

A thermal delay switch is usually mounted in a glass envelope and looks like a valve. The base-pins carry connections to the heater element and to the switch contacts.

A simple circuit which includes a thermal delay switch to protect a mercury vapour e.h.t. rectifier valve is shown in Fig. 10. V_1 is a normal hard-valve full-wave rectifier producing h.t. for the transmitter valves, and V_2 a mercury vapour e.h.t. rectifier. When S is closed, heater current is applied to V_1 and V_2 and also to the thermal delay switch. V_1 produces the h.t. output but relay contact A_1 prevents the a.c. being applied to V_2 anode. The winding of relay A/1 is connected to the h.t. output via a current limiting resistor R and the contacts of the thermal delay

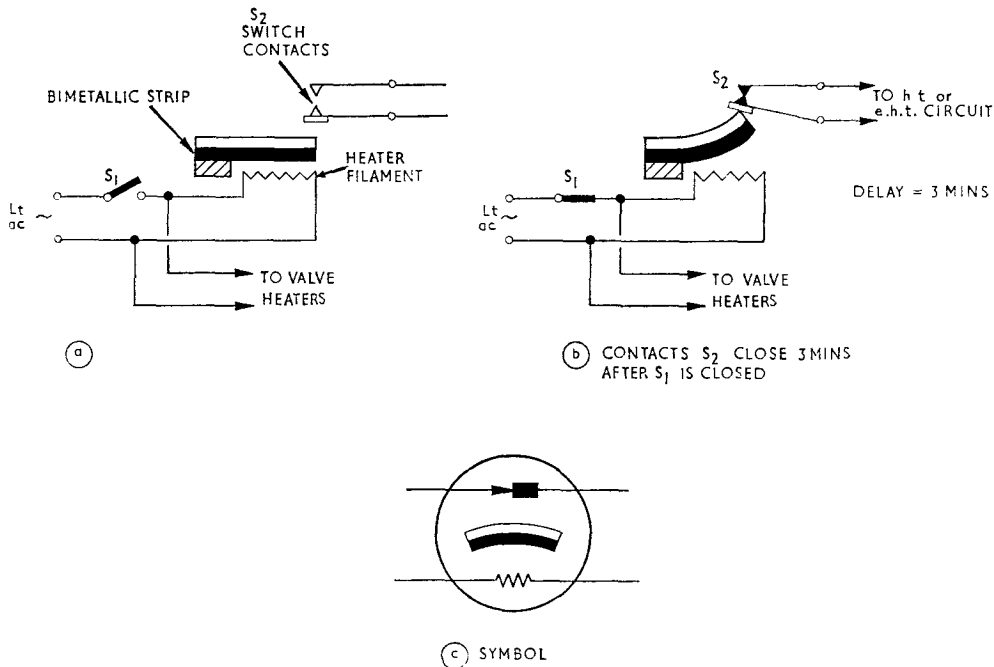


FIG. 9. THERMAL DELAY SWITCH

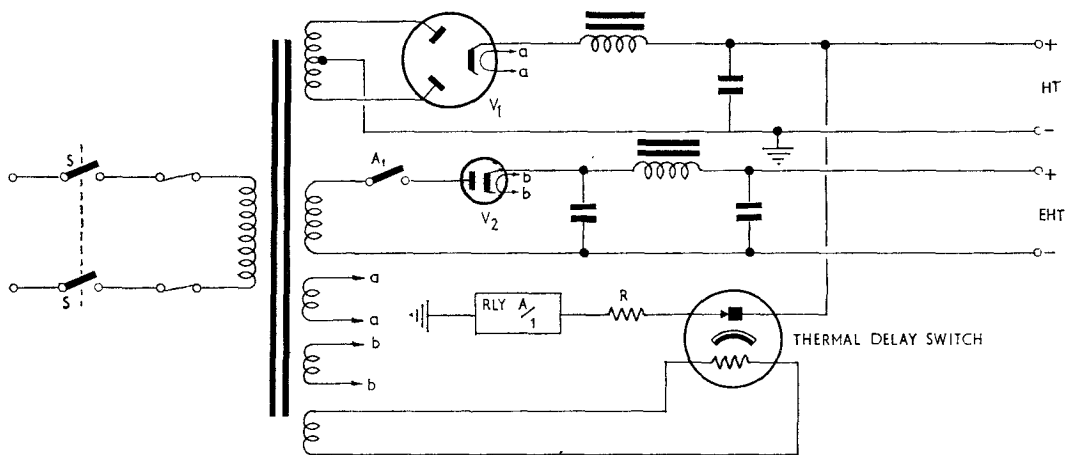


FIG. 10. SIMPLE DELAY CIRCUIT

switch. After a delay set by the delay switch which allows sufficient time for the mercury in V_2 to vaporise, the relay winding is energised and $A1_1$ contacts close, applying a.c. to the anode of the mercury vapour rectifier valve.

The delay imposed by a thermal delay switch can be increased by connecting two or more delay switches in series as shown in Fig. 11. The contacts on delay switch 1 close one minute after closing switch S . These contacts then complete the heater circuit for delay switch 2. After a further one minute, delay switch 2 contacts close and complete the required circuit.

Voltage Stabilisation

Variations in voltages applied to the transmitter can cause changes in radiated frequency and p.r.f. This is undesirable in most radars and special voltage stabilisation circuits are used to maintain the transmitter voltages at the designed values. The a.c. input to the rectifier and the d.c. output of the rectifier can both be stabilised.

AC stabilisation. One method of regulating the a.c. input to a power unit is with an *induction regulator*. This is a variable transformer, the primary winding of which can be moved so that the coupling between it and the secondary is variable (Fig. 12).

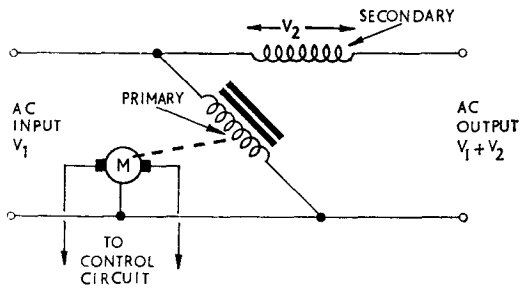


FIG. 12. BASIC INDUCTION REGULATOR

The a.c. input voltage V_1 is developed across the primary winding; the output voltage is the sum of V_1 and the secondary voltage V_2 . When the primary winding is at right angles to the secondary, the secondary voltage V_2 is zero. Thus the a.c. output equals V_1 , the input voltage. As the primary is turned into line with the secondary, V_2 gradually increases and the output voltage, $V_1 + V_2$ rises.

To make the induction regulator automatic the motor M must turn whenever the output voltage is above or below that required. A suitable motor control circuit is shown in Fig. 13.

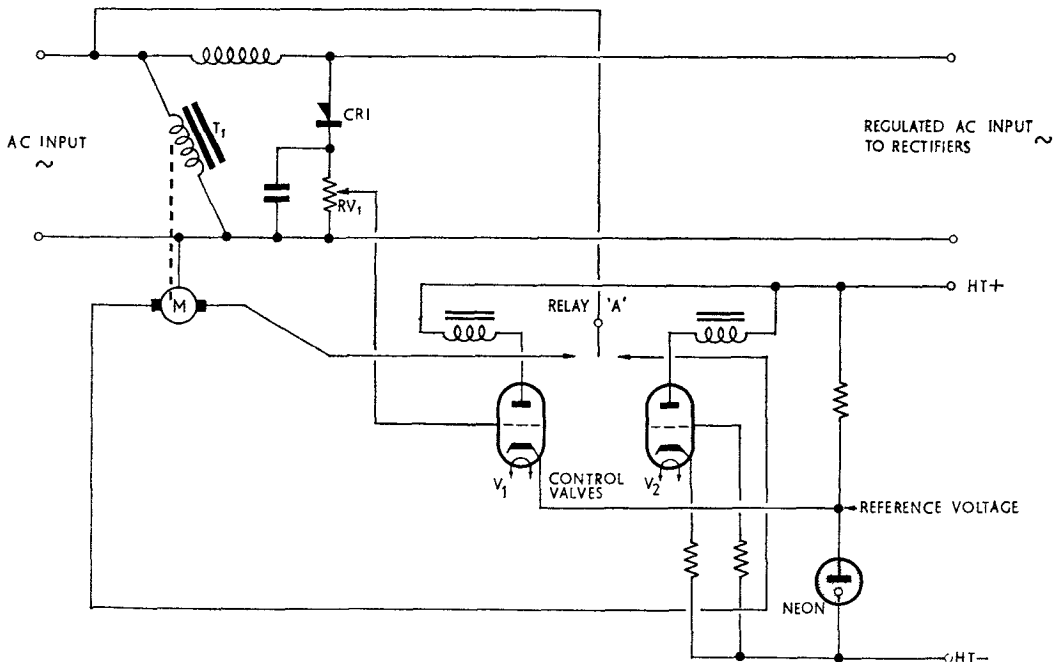


FIG. 13. INDUCTION REGULATOR WITH CONTROL CIRCUIT

The a.c. input is applied to the primary winding of the induction regulator T_1 . The output voltage is applied to a half-wave rectifier CR1, the rectified output from which is fed as bias to the grid of control valve V_1 . Potentiometer RV_1 is adjusted so that V_1 and V_2 conduct equally when the a.c. output is at the desired value. If the a.c. input varies, V_1 bias alters and balanced relay A energises causing the control motor to turn until currents through V_1 and V_2 are again equal.

Other types of a.c. voltage regulator include an autotransformer where the position of the wiper arm and hence the output voltage is controlled by a motor (Fig. 14a). The automatic motor control circuit may be similar to that of Fig. 13.

A saturable reactor can be used as an a.c. regulator as shown in Fig. 14b. If the input voltage varies, the current through the control winding varies and the change in reactance of the

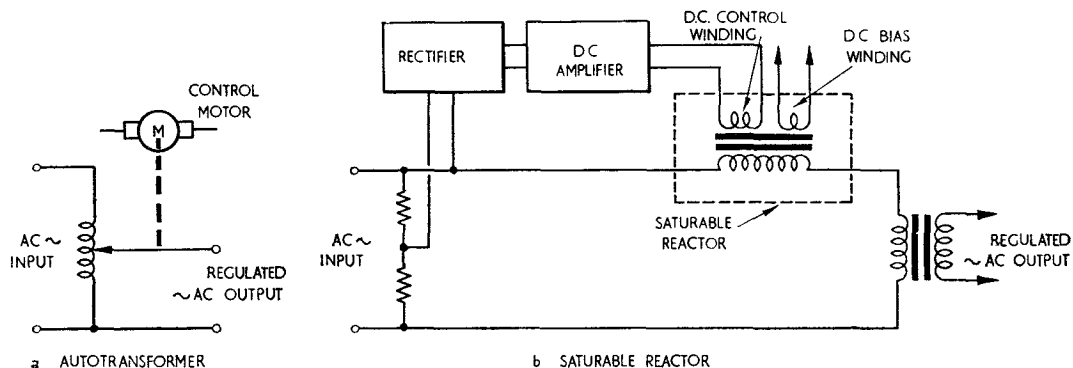


FIG. 14. OTHER TYPES OF A.C. VOLTAGE REGULATOR

the saturable reactor maintains the output constant. A big advantage of the saturable reactor is that since there are no moving parts it needs little maintenance.

DC stabilisation. If the load current from the rectifier varies, the output voltage will change. To reduce this effect most radar power units include *d.c. stabilisation* circuits. As discussed in Part 1 of these notes neon stabilisers (Fig. 15a) and zener diodes (Fig. 15b) can be used to maintain the load voltage constant.

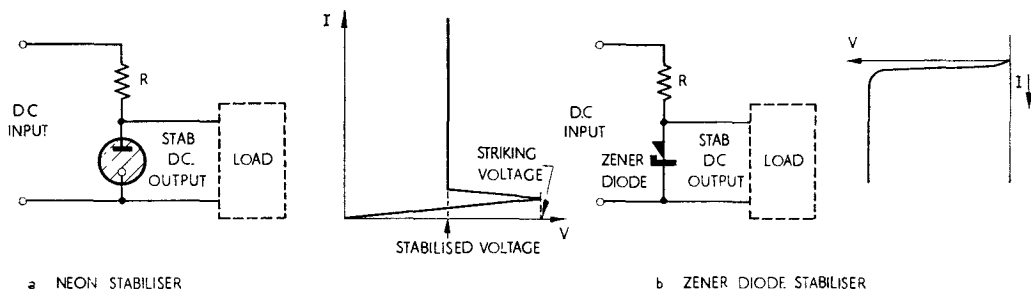


FIG. 15. NEON AND ZENER DIODE STABILISERS

Hard-valve stabilisation circuits are often used when the rectifier provides high voltages and currents. Fig. 16a shows a typical circuit using a high-current beam tetrode (V_1) as the series valve, and a pentode control valve (V_2). A neon stabiliser (V_3) holds V_2 cathode at a reference voltage set by the voltage divider network R_1 and R_2 .

If the load current increases, the voltage dropped across V_1 increases and the output voltage tends to decrease. The voltage across RV_1 falls, increasing the bias on V_2 . V_2 anode voltage

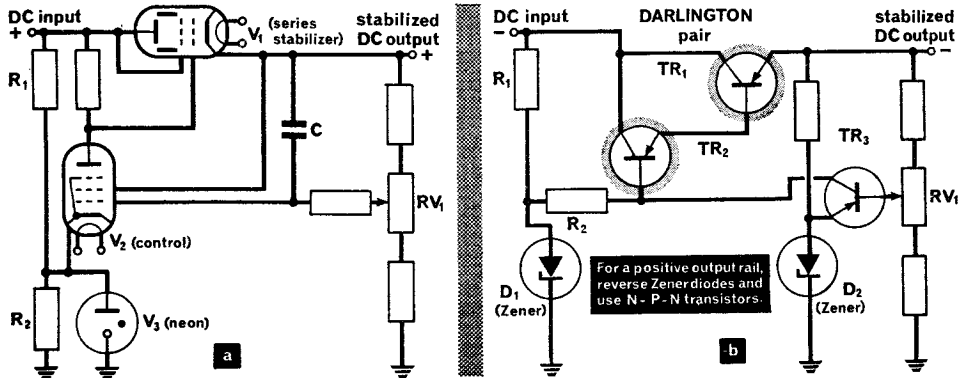


FIG 16. HARD VALVE AND TRANSISTOR SERIES STABILISERS

risers causing V_1 grid voltage to rise thus decreasing the series conducting resistance of V_1 and decreasing the voltage dropped across V_1 . This counteracts the initial fall in output voltage. Because of the high gain of V_2 small changes in output are rapidly corrected. Sudden changes are coupled directly to V_2 grid via C thus ensuring a more effective action. RV_1 enables the output voltage to be adjusted to the desired level.

A simple stabilisation circuit using transistors is shown in Fig. 16b. TR_3 emitter voltage is held at the reference level set by D_2 . Thus, if the output voltage tends to rise, TR_3 base-emitter voltage rises via RV_1 and this causes TR_3 collector current to rise. This rise in current through R_2 causes a fall in voltage at the base of TR_2 which therefore draws less current through the base emitter circuit of TR_1 . This causes the series resistance of TR_1 to rise, thereby reducing the output voltage to compensate for the initial tendency to rise. The compound-connected transistors TR_1 and TR_2 (usually known as darlington pair) are equivalent to a single transistor of very high current gain. This arrangement is very common in practical power amplifier circuits, particularly in d.c. amplifiers.

Thyristors

A thyristor, also known as a silicon controlled rectifier, is a semiconductor device used in a.c. power control circuits, d.c. converters and in voltage-regulated d.c. supply circuits. It consists of four layers of p-n-p-n semiconductor and has three terminals, an *anode*, a *cathode* and a *gate* or *trigger* electrode (Fig. 17a).

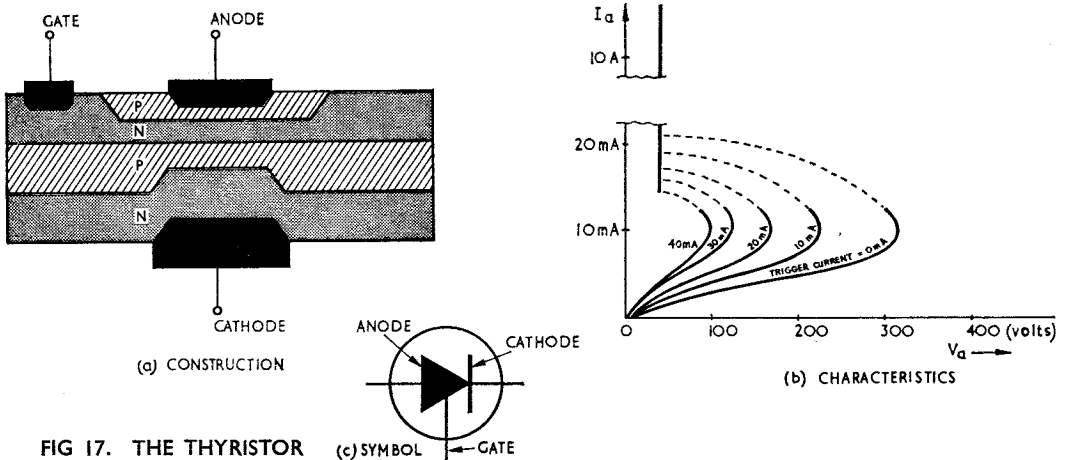


FIG 17. THE THYRISTOR

When an alternating voltage is applied between anode and cathode no current flows. However, if a small current is fed to the gate terminal when the anode is positive relative to the cathode the thyristor switches rapidly into the conducting state and behaves as a low-resistance diode with a forward resistance of approximately 0.001 ohms. Current continues to flow even when the trigger current is removed, until the anode voltage has fallen to a low value. From Fig. 17b it can be seen that the characteristics of a thyristor are very similar to those of a thyatron.

With a trigger power of only 5V at 100 mA the thyristor can pass a current of 50A at 250V. Thus, assuming a forward resistance of 0.001 ohm the voltage dropped across the thyristor is only 0.05 volt, the remaining voltage being developed across the load. The power dissipated in the thyristor is therefore only 2.5 watts. With no trigger input an applied voltage of almost 350V is required to produce breakdown. The thyristor can be triggered by very short-duration pulses applied to the trigger electrode. When breakdown occurs, current through the thyristor builds up to a maximum in a few microseconds. Thus, with fairly low power applied to the control gate, the thyristor can *control* very large powers fed to a circuit.

Thyristor as a Stabiliser

The two main disadvantages of using a series valve or transistor as a stabiliser is that the regulation is not effective over a very wide range of voltage, and considerable power is wasted in the series valve, i.e. the efficiency is low.

An alternative method of regulation is to switch the series regulator on for only a part of the conducting half-cycle. This can be done by using a thyristor as the series regulator and timing the triggering pulse to occur at precise instants during the positive half cycle of input. By varying the instant (t_1 , t_2 , t_3 in Fig. 18) at which the thyristor is switched on the mean level of output voltage from the smoothing filter can be varied. The principle is similar to that of the phase sensitive rectifier and is illustrated in Fig. 18.

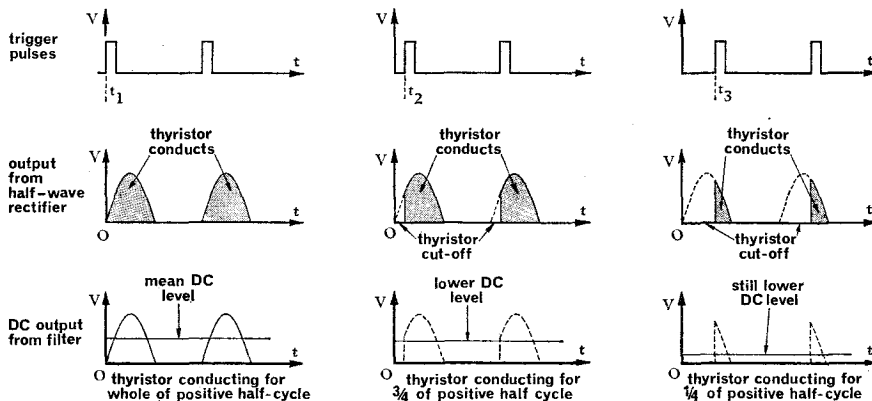


FIG 18. PRINCIPLE OF SWITCHED REGULATION

If we arrange that any small change in the desired d.c. output voltage is amplified and used to alter the instant at which the series thyristor fires, the voltage can then be brought back to the desired level and we have an automatic voltage stabiliser circuit. The block diagram of a suitable circuit is shown in Fig. 19. If the output voltage changes, an error signal is applied to the control circuit which alters the timing of the firing pulse produced by the pulse generator. The p.r.f. of the firing pulse is synchronised to twice the frequency of the a.c. supply, i.e. one pulse occurs every half-cycle of full-wave rectification.

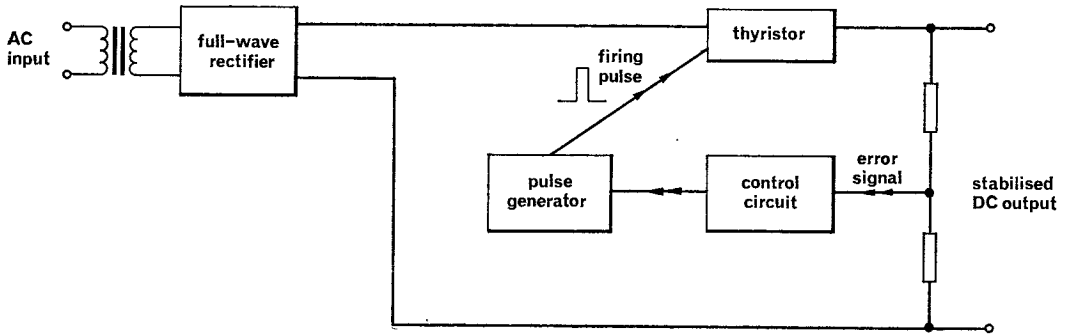


FIG 19. OUTLINE OF THYRISTOR SWITCHED VOLTAGE STABILISER CIRCUIT

Switched Rectifier Circuit using Thyristors

Many modern power supply circuits use thyristors as switched rectifiers and have a circuit arrangement similar to that shown in Fig. 20. Thyristors used in this way are normally called silicon controlled rectifiers (SCRs).

The two SCRs are used instead of the diodes of a conventional full wave rectifier/smoothing circuit, but each one only conducts for part of the positive half cycle and therefore can supply a variable amount of power to the smoothing circuit. The switch-on point is controlled by a delayed positive half cycle applied to the gate from T_2 . This delay is varied by adjusting the phasing of the coupling circuit between T_1 and T_2 by varying the effective conducting resistance of the transistors TR_1 and TR_2 . These transistors (one for each half cycle) are controlled by the output of a d.c. amplifier which senses changes in the voltage level of the rectified output. The phasing is so arranged that if the output tends to rise; the SCRs are switched on later in the half cycle, provide less power to the smoothing circuit, causing the output to fall, thus compensating for the initial tendency to rise.

A circuit of this type is capable of giving a regulated output to an accuracy of about 1% at high powers (several kW). The output voltage can be adjusted to the required level 'on-load' by adjusting RV_1 .

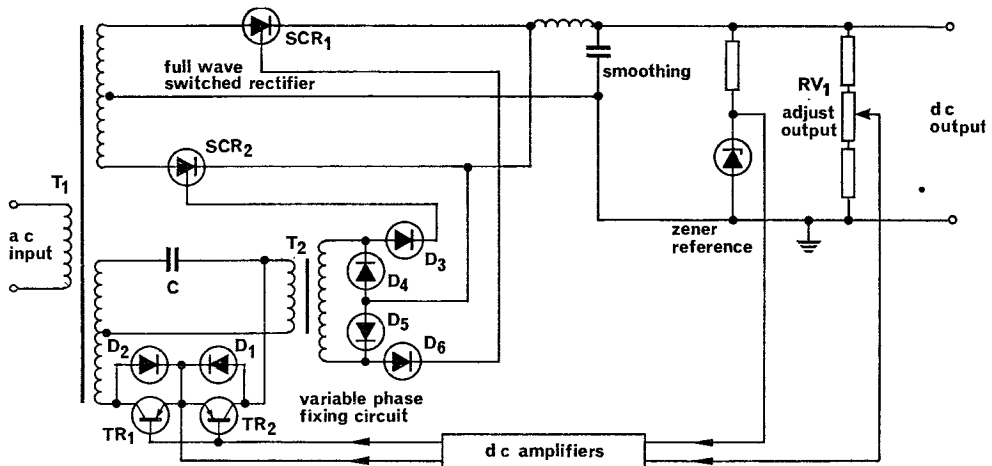


FIG 20. THYRISTOR RECTIFIER AND STABILISER

Interlock and Protection Circuits

Because high voltages and currents are produced in a radar transmitter, interlock and protection systems are always used to protect components and circuits as well as to safeguard against electric shock.

Fig. 21 shows a simple interlock system which is designed to prevent the e.h.t. voltage being applied to the modulator, and to prevent the sub-modulator valve operating unless the modulator and transmitter doors or panels are correctly closed, the magnetron heater has been on for three minutes and the magnetron blower motor supplies are complete.

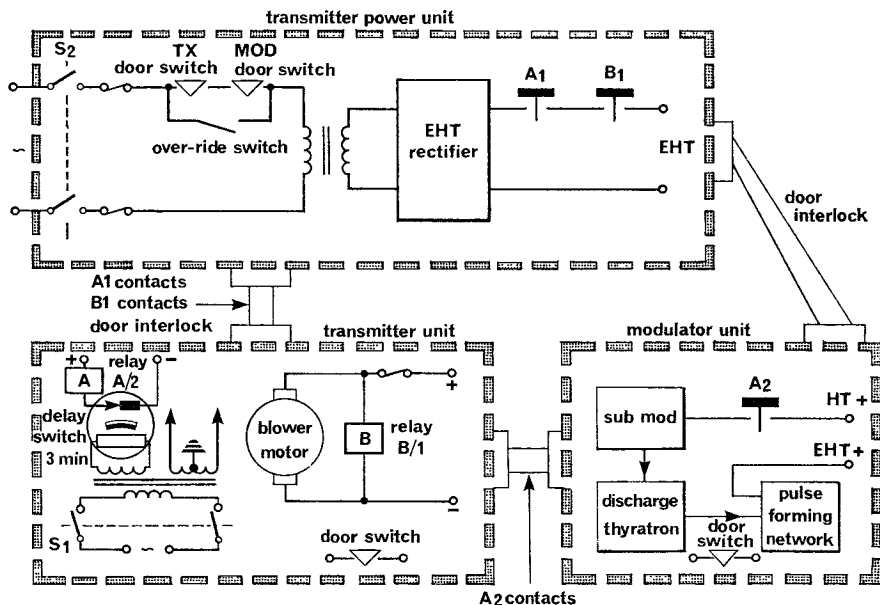


FIG 21. SIMPLE INTERLOCK AND PROTECTION SYSTEM

When the transmitter switch S₁ is closed current is applied to the magnetron heater and to the thermal delay heater. After a delay of three minutes relay A/2 energises, closing contacts A₁ in the e.h.t. power unit and A₂ in the modulator unit. If the supply to the magnetron blower motor is complete (relay B/1 energised) and the doors on the transmitter and modulator units are closed, the e.h.t. and sub-modulator supplies are completed and the transmitter will operate.

If the units are not correctly connected together or the heater and blower motor supplies are inoperative, or either of the two doors is open, the e.h.t. modulating pulse will not be applied to the magnetron. Sometimes an over-ride switch is placed in parallel with the door switches to enable the modulator and transmitter units to be serviced with the doors open and e.h.t. applied.

SECTION 6
RADAR RECEIVERS

Chapter		Page
1	Basic Requirements of a Radar Receiver	395
2	Stages in a Radar Receiver	407
3	Reduction of Clutter	427
4	Interference and Jamming	437

CHAPTER 1

BASIC REQUIREMENTS OF A RADAR RECEIVER

Introduction

In Section 1 (p. 23) we built up the schematic diagram of a basic primary radar installation, and as we have progressed we have investigated each block in turn. The only block which we have not yet looked at in detail is the *receiver* (Fig. 1).

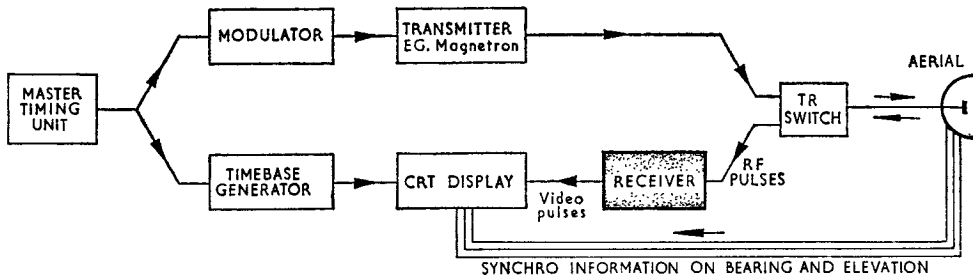


FIG. 1. BASIC PRIMARY RADAR BLOCK DIAGRAM

The purpose of a radar receiver is to accept target echoes in the form of r.f. pulses and from them produce an output in the form of video pulses for application to the c.r.t. display. The video pulses must be undistorted, relatively free from noise and of sufficient amplitude to produce an indication of the target on the c.r.t. screen.

The r.f. input pulses to a radar receiver are of very short duration (of the order of microseconds) and of very small amplitude (of the order of microvolts). The characteristics of a receiver capable of handling this type of signal input are therefore different from those required by a communications receiver. In a communications system the input is usually much stronger, since the input energy is from a transmitter and not due to *reflection* as in primary radar. In addition, the signal is usually some form of c.w., either modulated or unmodulated.

In this chapter we shall discuss the special characteristics of a radar receiver. Briefly, the desired characteristics are:

- High gain* so that the weakest echo may be detected.
- Low receiver noise factor* so that the signal-to-noise ratio is kept as high as possible.
- Sufficiently wide bandwidth* to prevent distortion of the received pulse.

These characteristics are more easily obtained in the superheterodyne receiver than in any other type and, because of this, most radars use a superhet. It is the only type of radar receiver we shall discuss in these notes.

Radar Receiver Block Diagram

A simple block diagram of a radar superheterodyne receiver is shown in Fig. 2. With a few exceptions this diagram looks very similar to that of the basic superhet discussed in Part 1B of these notes (p. 406). The differences are in the design and detailed circuitry of the receiver and these are discussed in more detail later.

- RF amplifier.** The echo signal picked up by the aerial passes through the TR switch and is amplified by the low-noise r.f. amplifier. Until recently, r.f. amplification was difficult at microwave frequencies and in many practical centimetric receivers the r.f. amplifier stage is left out and the mixer acts as the first stage.

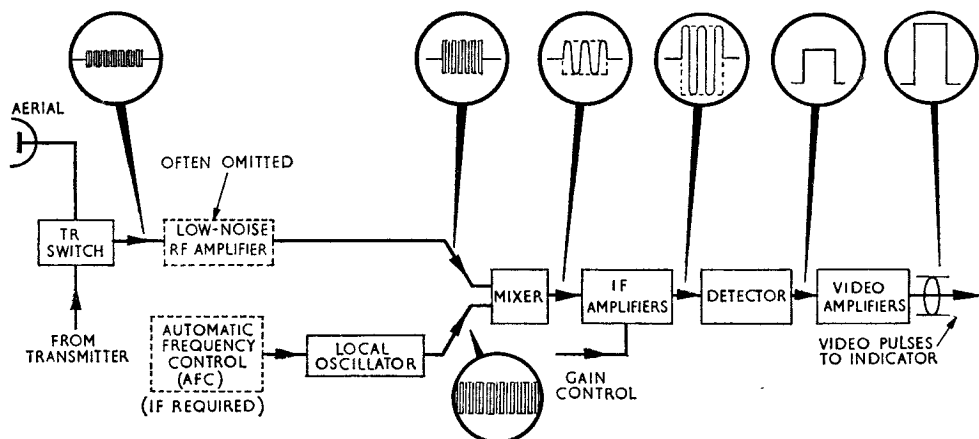


FIG. 2. BLOCK DIAGRAM OF RADAR SUPERHETERODYNE RECEIVER

b. Mixer. At the mixer, the echo signal beats with the output of a local oscillator to produce an output at the i.f. Additive mixing is used. At frequencies up to about 1,000 Mc/s, planar (disc-seal) triodes may be used both for the mixer and the local oscillator. Above this, the mixer uses a silicon or germanium crystal diode, with a reflex klystron as the local oscillator. Values of i.f. in radar receivers are usually within the band 10 Mc/s to 60 Mc/s. Intermediate frequencies that have been used in Service radar receivers include 13.5 Mc/s, 30 Mc/s and 45 Mc/s.

c. IF amplifiers. The output from the mixer is amplified by the i.f. amplifiers. It is in these stages that most of the receiver gain is obtained and, to achieve a high gain, six or more i.f. stages are normally used, each carefully designed to provide stability. Very often the first two i.f. stages are designed as low-noise amplifiers and are placed, together with the TR switch, r.f. amplifier (if fitted), mixer and local oscillator, in the 'r.f. head' of the radar equipment (Fig. 3). The r.f. head is situated close to the aerial so that loss of signal through connecting waveguide or coaxial cable is negligible. The i.f. output from the r.f. head is then taken via coaxial cable to the main i.f. amplifier in the receiver unit; at this lower frequency losses in the connecting cable are small. The wide bandwidth necessary to preserve the shape of the received pulse is also determined by correct design of the i.f. amplifiers.

d. Detector. The amplified i.f. output is applied to the detector. This is a conventional amplitude detector using either thermionic or semiconductor diodes. The output may be either positive-going or negative-going as required, depending upon the detector connections.

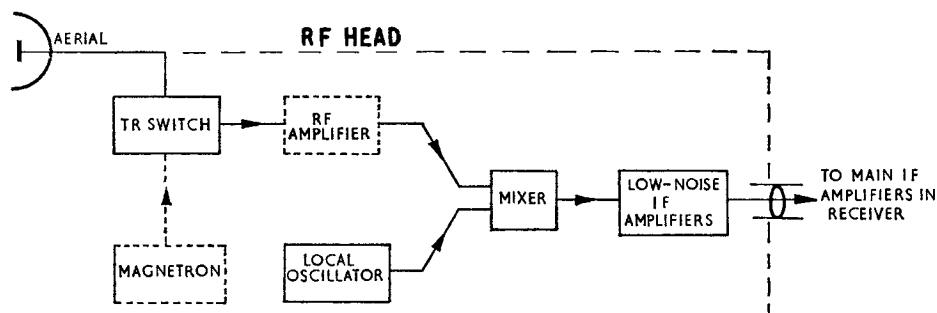


FIG. 3. USE OF RF HEAD

e. Video amplifiers. The detector output is amplified by the video amplifiers. About two stages are normal. These are designed to have the required bandwidth. In addition they usually provide amplitude limiting of large signals to prevent overloading the display. The final video amplifier is usually connected as a cathode follower to provide a low source impedance for correct matching into the coaxial cable feeding the video signals to the c.r.t. indicator. The video pulses are applied either to the grid (or cathode) for intensity-modulated displays such as the p.p.i., or to the Y deflector plates for deflection-modulated displays such as the type A.

f. AFC. Most centimetric radars use a magnetron as the transmitter and a reflex klystron as the local oscillator in the receiver. The magnetron suffers considerably from frequency drift and to keep the output from the signal mixer in the receiver at the correct i.f. we must arrange for the klystron local oscillator to shift in frequency by the same amount as the magnetron drifts. This is the job of the automatic frequency control (a.f.c.) circuit which we shall discuss in more detail later.

Having run through the various stages in a radar receiver let us now look more closely at the problems to be overcome in providing the required receiver characteristics.

Gain

From previous work we know that the voltage induced in a radar aerial by an echo signal may be extremely small, frequently as small as $1\mu\text{V}$. The c.r.t. indicator, on the other hand, may require a signal deflection voltage of 10V or more to operate satisfactorily. Thus we may require a voltage gain, from the aerial to the indicator, of $\frac{10}{10^{-6}} = 10^7$. This is equivalent to a power gain of 140 db, and is typical of the required gain of a radar receiver; the radar type 80 receiver, for example, has a gain of 125 db.

Most of this gain occurs in the i.f. amplifiers. An r.f. amplifier, if used, might provide a gain of 10 (20db), the video amplifiers a gain of 10 (20 db) and the i.f. amplifiers a gain of 100,000 (100 db). The i.f. amplifiers must also provide a bandwidth sufficient to pass the pulse without distortion. For any single stage, wide bandwidth generally leads to low gain. Hence to provide the necessary gain together with wide bandwidth, a large number of i.f. amplifier stages are needed. Most radar receivers use at least six stages; some receivers have as many as ten. In amplifiers with such a high gain, special precautions are taken to prevent instability. These precautions include careful screening, adequate decoupling and the correct application of negative feedback.

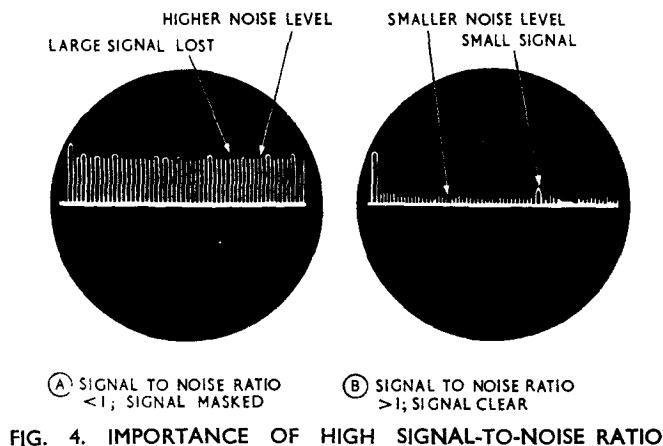
Noise

It may appear from the previous paragraph that by providing the receiver with sufficient gain it would be possible to detect any signal, no matter how weak. This would be true were it not for *noise*. We saw on p. 20 that noise can enter the receiver via the aerial along with the desired signal (external noise) or it can be generated within the receiver itself (internal noise). No matter how the noise is produced, its effect on the c.r.t. display is the same: in a type A display noise appears as 'grass'; in a p.p.i. display it appears as 'snow'; in both cases, if the noise level is high, it may completely obscure the desired signal echo.

Noise is always present in a receiver to some extent and it is amplified in the same proportion as the signal. Hence if the noise level is initially greater than that of the signal, subsequent amplification is *useless* because both are amplified by the same amount and, on the c.r.t., the desired signal is completely masked by the noise (Fig. 4a).

In a radar receiver, the minimum detectable signal depends upon the amount of noise present; the limit of sensitivity is reached when the signal level falls below that of noise. If there is a high noise level, the signal amplitude must be even higher if the target is to be seen on the c.r.t. To

detect a weak signal it is necessary for the noise level to be low (Fig. 4b). Thus the important factor in the detection of signals is the *signal-to-noise ratio*. The aim must always be to have as great a signal-to-noise ratio as possible by increasing the signal, or by reducing the noise, or both.



External Noise Sources

The receiver sensitivity may be limited by external noise which enters the receiver, via the aerial, along with the signal. The main sources of external noise include:

a. Man-made noise. This includes noise from car ignition, fluorescent lighting, electric generators and electric motors. The noise from such sources is negligible at frequencies above about 300 Mc/s.

b. Atmospheric noise. This is caused by electrical storms but produces very little noise above about 100 Mc/s.

c. Cosmic noise. A continuous background of noise, known as cosmic noise, is radiated from outer space. In general it decreases with increase in frequency and is only of importance at frequencies up to about 1,000 Mc/s.

d. Atmospheric absorption noise. The atmosphere absorbs energy from a radar wave (atmospheric attenuation). Some of this absorbed energy is re-radiated by the atmosphere, and the re-radiation takes the form of noise. Absorption noise is important at frequencies above about 10,000 Mc/s (3 cm). This coincides with the fact that atmospheric attenuation becomes important above 10,000 Mc/s (see p 228).

At microwave frequencies the external noise level is low in relation to internal (receiver) noise, and for radar receivers using a mixer as the first stage the sensitivity is determined mainly by the noise generated within the receiver itself. However, some modern microwave receivers use low-noise r.f. amplifiers as the first stage (e.g. parametric amplifiers and masers) and the sensitivity of such receivers may be limited more by external noise than by internal receiver noise. There is no cure for external noise since most of it comes from sources outside man's control. Thus even if it were possible to have a perfect receiver, which produces no noise of its own, the external noise sets a limit to the magnitude of the signal which can be detected.

Receiver Noise

The main sources of internal noise are discussed below.

a. Thermal noise. This is sometimes called *Johnson noise*. In any conductor at a temperature other than absolute zero there is a random movement of electrons. This movement

of electrons constitutes a current which produces a random noise voltage across the conductor. The magnitude of the noise voltage depends upon three factors: the *temperature* of the conductor, its *resistance*, and the *bandwidth* of the circuit.

This noise may be expressed as an r.m.s. voltage e_n :

$$e_n = \sqrt{4kTBR} \text{ volts,}$$

where k is a constant (Boltzmann's constant),

T is temperature in degrees Kelvin,

B is the circuit bandwidth in c/s,

R is the resistance in ohms.

The equivalent noise source of a radar aerial, due solely to thermal noise in the aerial system, can be considered to consist of a noise generator of e_n volts in series with a noise-free resistance R equal to the aerial resistance. The maximum power transfer theorem states that maximum power is transferred when the load is correctly matched to the source, i.e. when $R_{\text{load}} = R_{\text{source}}$. Thus when the aerial is correctly matched to the receiver input circuit, Fig. 5 applies, and the maximum noise power transferred from the aerial to the receiver is:

$$\frac{e_n^2}{4R} = \frac{4kTBR}{4R} = kTB \text{ watts.}$$

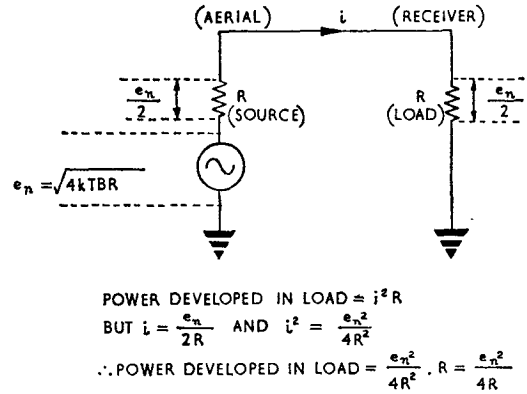


FIG. 5. MAXIMUM AVAILABLE NOISE POWER

This is the *maximum available* noise power at the receiver input due to thermal noise at the input. It is also the condition for maximum transfer of available *signal* input; so the maximum available signal-to-noise ratio at the input is $\frac{P_s}{kTB}$, where P_s is the signal power.

To detect a signal, the signal must be at least as powerful as the noise; hence the minimum detectable signal power is also kTB . In practice, for anything useful, the signal-to-noise ratio $\frac{P_s}{kTB}$ must be greater than 1, and P_s must *exceed* kTB .

The power spectrum of thermal noise is *uniform* throughout the radio frequency bands. A 2 Mc/s band produces the same amount of thermal noise anywhere within the radio spectrum. Noise of this nature is termed *white noise*. Note particularly that the thermal noise power input is proportional to the *bandwidth* of the circuit. If the circuit passes a wide band of frequencies, the noise power is large. By reducing the bandwidth we reduce the noise power developed in the circuit.

b. Shot noise. Electrons are emitted at random from the cathode of a valve and reach the anode at a *non-uniform* rate. This gives rise to a variation of current which produces a noise voltage across the load resistor. In effect, the valve current is a direct current whose amplitude fluctuates in a random manner, i.e. modulated by noise.

When the valve is operating under space-charge conditions, the space charge tends to smooth out the random variations in current and shot noise is reduced. A saturated valve, on the other hand, produces considerable shot noise (a saturated diode is often used as a noise generator).

Shot noise occurs in all hot-cathode valves, in semiconductor diodes, transistors and other devices which carry current.

c. Partition noise. In a valve with more than one positive electrode the current divides between the electrodes. Since the distribution of emitted electrons is random, the *division* of current between the various positive electrodes is also random and *additional* noise is introduced. The more collecting electrodes there are, the more variation there is in the *division* of the original electron stream and the greater is the *partition noise*. A pentode is about five times as noisy as a triode, and a heptode about twenty-five times as noisy, for these reasons. Thus, when low noise is desired, as in the input stages of a radar i.f. amplifier, triodes are preferred.

d. Induced grid noise. At high frequencies, when the transit time of electrons across a valve becomes significant, the number of electrons approaching the grid at a given instant may not equal the number receding from the grid. A voltage is therefore induced in the grid. In addition, because the emission of electrons is random (shot noise), the voltage induced in the grid is also random and induced grid noise is produced. This form of noise increases with increase in frequency because of the increasing transit time effect.

e. Semiconductor noise. Semiconductor diodes and transistors are subject to shot noise and thermal noise in the same way as valves. Semiconductor devices also suffer from a form of low-frequency noise, known as *flicker noise*. Flicker noise increases with *decrease* of frequency and can be neglected above about 100kc/s. It is, however, important in second detector and video stages.

Noise Factor

The term 'signal-to-noise ratio' of a receiver is generally taken to mean the ratio of the output signal power to the output noise power. With an ideal receiver, i.e. one that generates no internal noise, the signal-to-noise ratio at the output of the receiver would be the *same* as that at the input. In a practical receiver, as we have seen, each stage adds to the thermal noise present at the aerial and this is amplified and added to by succeeding stages. Hence the *output* signal-to-noise ratio is much *lower* than the *input* signal-to-noise ratio.

If the input signal power is $1\mu\text{W}$ and the input noise power $0\cdot01\mu\text{W}$, the input signal-to-noise ratio is 100:1 (20db). The *output* signal-to-noise ratio is less than this. Fig. 6 shows

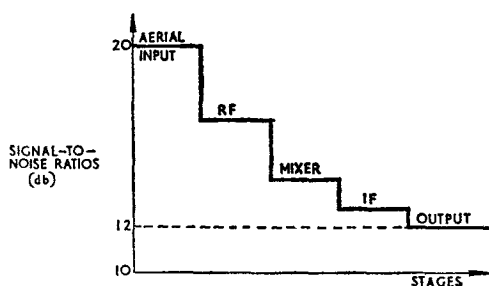


FIG. 6. SIGNAL-TO-NOISE RATIO AT VARIOUS STAGES IN A RECEIVER

typical signal-to-noise ratios in db at the various stages of a receiver. In Fig. 6 the output signal-to-noise ratio is 12 db (a power ratio of 16:1). Thus if the amplification of the signal and the noise is the same, the receiver increases the input noise by $\frac{100}{16}$ or 6.25 times.

The *ratio* of the signal-to-noise ratio at the *input* of a receiver to that at the *output* is known as the *noise factor* (or noise figure) of the receiver. It is a measure of the noise introduced by the receiver itself. We can therefore write:

$$\text{Noise Factor } N = \frac{\text{Input signal-to-noise ratio}}{\text{Output signal-to-noise ratio}}$$

If the receiver were perfect, in that it contributed no noise of its own, the signal-to-noise ratio at the output would be the same as that at the input, in which case N would be 1 (0 db). For all practical receivers N is greater than 1. For the receiver of Fig. 6:

$$N = \frac{100:1}{16:1} = 6.25.$$

Expressed in terms of decibels (i.e. $10 \log_{10} N$) the noise factor of the receiver in Fig. 6 is 8 db. It is the *difference* between the input and output signal-to-noise ratios measured in db (i.e. 20 db — 12 db). The figure of 8 db is typical of the noise factor of radar receivers.

Table 1 shows the advantage of using decibels.

	Ratios	Decibels
Input signal-to-noise	100:1	20 db
Output signal-to-noise	16:1	12 db
Noise factor	$\frac{100:1}{16:1} = 6.25$ (A complicated division)	$20 - 12 = 8$ db (A simple subtraction)

TABLE 1. NOISE FACTOR EXPRESSED IN RATIOS AND IN DECIBELS

The aim of the designer is to produce a receiver with as low a noise factor as possible. This is complicated by the fact that the receiver has to accept very narrow pulses and hence a *wide band* of frequencies. Wide bandwidth increases the receiver noise factor because the available thermal noise power is kTB , where B is the bandwidth of the circuit. Thus the design of a receiver is a compromise between high gain, wide bandwidth and low noise factor.

Importance of Low Noise Factor in First Stage

So far we have considered noise factor as applied to the whole receiver. However, the idea of noise factor can also be applied to a single stage or to several stages in cascade. The *first stage* in a receiver is the most important in determining the overall noise factor of the receiver because its noise is amplified by all the succeeding stages. The aim is to provide a first stage with as *high a gain* and as *low a noise factor* as possible. Note that high gain and low noise factor are *both* required. If the first stage has a low noise factor and a *low gain* the noise factor of the *second* stage becomes important. Thus in those microwave receivers which do not use an r.f. amplifier, the first stage (mixer) has a gain less than 1. Consequently, the succeeding i.f. amplifier must be designed with a high gain and a low noise factor. The use of modern r.f. amplifiers before the mixer improves this situation.

Suppose we have a receiver in which the noise factor of the mixer is 16 (12 db). By inserting an r.f. amplifier, which has noise factor of 2 and a gain of 10, the overall noise factor is reduced to 3.5 (5.4 db). It is therefore primarily the *first stage* in a receiver which determines the noise factor of the receiver as a whole (Fig. 7).

Measurement of Noise Factor

We have previously defined noise factor as the signal-to-noise ratio at the input of a receiver divided by the signal-to-noise ratio at its output. We can define it another way. The noise generated within the receiver, together with thermal noise from the aerial, may be regarded as being equivalent to a particular noise input signal applied to an *ideal* receiver (i.e. one that

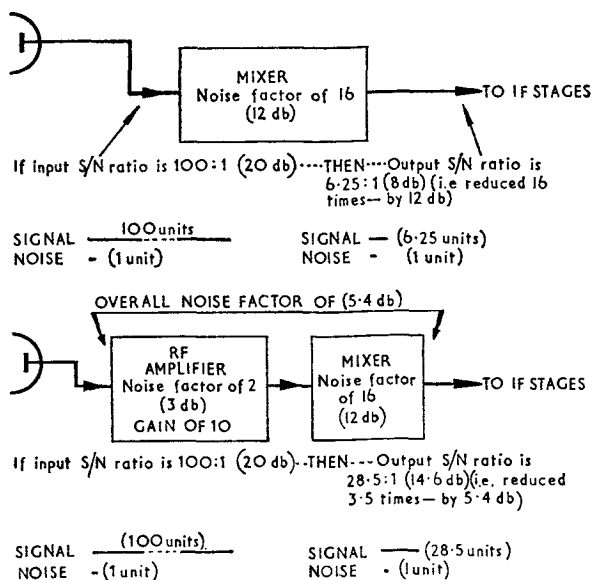


FIG. 7. REDUCTION OF NOISE FACTOR BY USE OF RF AMPLIFIER STAGE

produces no internal noise); the noise factor is then the ratio between this equivalent noise input signal and the thermal noise (Fig. 8). This idea is used directly in practical measurement of noise factor.

To measure noise factor we need a suitable noise generator and an output-power meter. The noise generator is connected to the aerial terminal of the receiver and the output-power meter is connected to the receiver output. Since the noise generator takes the place of the aerial with which the receiver is normally used, its output impedance must match that of the aerial.

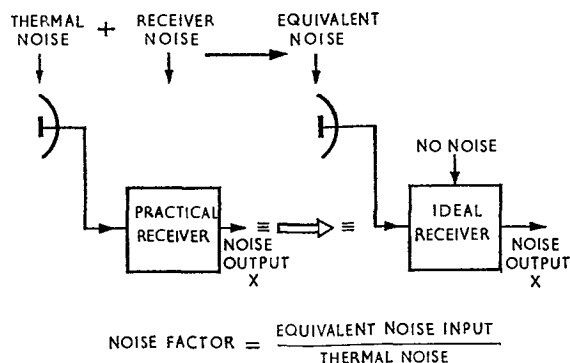


FIG. 8. MEASUREMENT OF NOISE FACTOR

We know that a saturated diode produces a high level of shot noise output. It is therefore suitable as a noise generator at frequencies up to about 1,000 Mc/s. A typical circuit is shown in Fig. 9. The h.t. voltage is sufficiently high to ensure that the diode remains saturated. The noise current through the valve then depends upon the temperature of the cathode and this is controlled as shown. The noise voltage is developed across the load resistor and matched into the receiver input.

With the noise generator connected, *but not switched on*, the noise power output of the re-

ceiver is measured on the output meter connected to the receiver output. The noise generator is then switched on and its noise level adjusted so that the output power of the receiver is *doubled*. At this setting the output level of the noise generator has the same value as the equivalent noise input level mentioned in Fig. 8. The noise generator is calibrated in terms of the ratio between its noise output and thermal noise so that it indicates the noise factor of the receiver directly.

At frequencies above about 1,000 Mc/s the noise generator is usually a gas discharge tube similar to the well-known fluorescent lighting tube. The tube is inserted into the waveguide as shown in Fig. 10. The noise output from the discharge tube is not variable (as it is from a diode) so a slightly different method of measurement of noise factor is necessary. With the noise tube

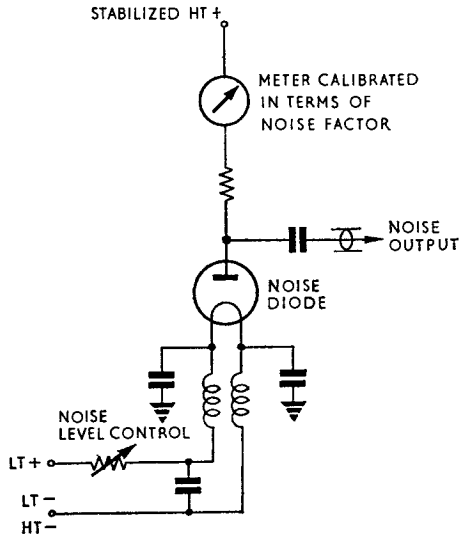


FIG. 9. DIODE AS A NOISE GENERATOR

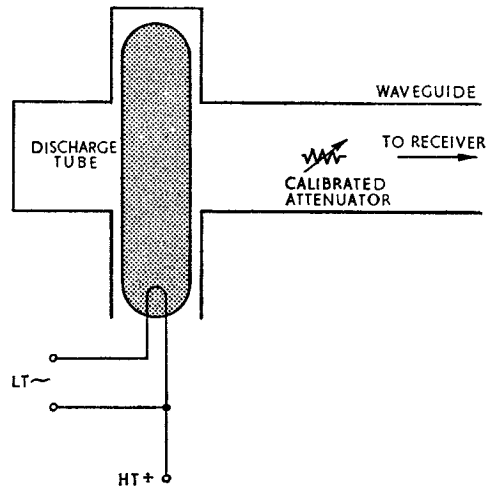


FIG. 10. GAS DISCHARGE TUBE AS A NOISE GENERATOR

switched off, the reading in the output-power meter is noted. The noise tube is then ionized and this causes the meter reading to increase. An attenuator in the waveguide feed from the noise tube to the receiver is then adjusted until the meter reading is the same as with the tube off. The noise factor is then read directly from the calibrated scale of the attenuator.

Typical noise factor readings for microwave radar receivers should be not greater than about 10 db. If a figure much higher than this is obtained, it is an indication that a fault exists in the system and this has caused the receiver sensitivity to decrease sharply.

Receiver Bandwidth

The required r.f. bandwidth of any receiver is largely determined by the type of signal applied to it. For example, an a.m. broadcast receiver deals with sound-modulated waves, the highest modulation frequency of which is about 8 kc/s. As the overall r.f. bandwidth should be *twice* the highest modulation frequency, the maximum required bandwidth for an a.m. broadcast receiver is about 16 kc/s.

The bandwidth required for a pulse-modulated radar signal is very much greater than that required for the sound-modulated signal. We have seen something of the reason for this in Section 2 (p. 64) where we showed that a rectangular pulse is made up of a sine wave of a certain fundamental frequency plus an *infinite* number of odd harmonics. Thus to accept a pulse-modulated signal and pass it *without distortion* we should require an infinite bandwidth! This, of course, is not possible, but every effort is made to pass as many as possible of the frequencies in the pulse, consistent with the introduction of as little noise as possible.

Fig. 11 compares the bandwidth required by a sine-wave modulated signal with that required by a pulse-modulated signal. If a pure audio frequency sine wave of frequency f_m is caused to amplitude-modulate an r.f. carrier of frequency f_o , only two side frequencies are produced (Fig. 11a). The upper side frequency occurs at $f_o + f_m$, whilst the lower side frequency is $f_o - f_m$. The required bandwidth is thus $2f_m$, typically 16 kc/s for sound-wave modulation as noted earlier.

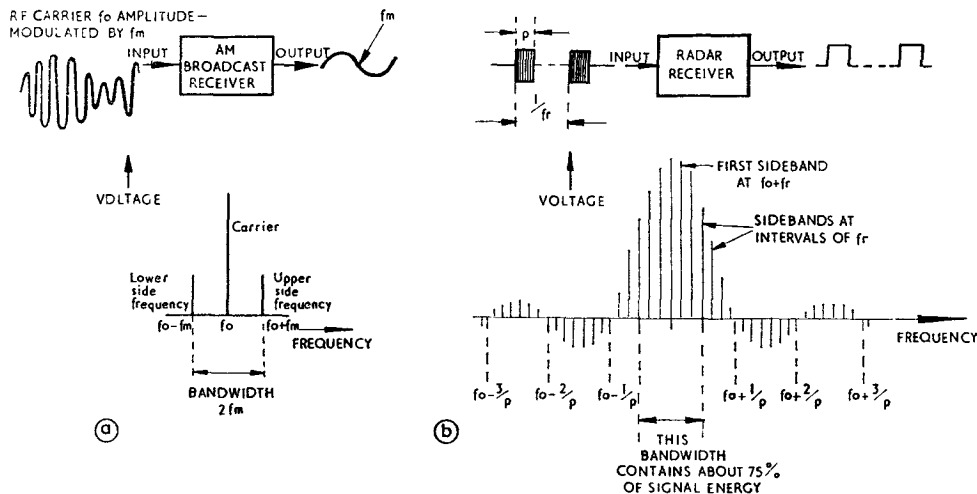


FIG. 11. BANDWIDTH REQUIREMENTS OF MODULATED SIGNALS

We noted earlier in these notes that a pulse consists of a sine wave of a certain fundamental frequency plus an infinite number of harmonics. Thus, when a wave train is pulse-modulated, as in radar, numerous side frequencies are produced. By applying Fourier's analysis to the r.f. pulses the spectrum shown in Fig. 11b is produced. The first pair of side frequencies occurs at $f_o \pm f_r$, where f_r is the pulse repetition frequency. But, in addition, side frequencies occur at intervals of f_r from the carrier frequency f_o , their amplitudes decreasing to zero at intervals of $\frac{1}{p}$ from f_o , where p is the pulse duration.

Examination of Fig. 11b shows that most of the energy is contained between the limits $f_o + \frac{1}{p}$ and $f_o - \frac{1}{p}$, corresponding to a bandwidth of $\frac{2}{p}$. In fact, about 75% of the energy is contained in half this bandwidth, i.e. a bandwidth of $\frac{1}{p}$. Thus a receiver bandwidth of between about $\frac{1}{p}$ and $\frac{2}{p}$ will pass most of the components in the received pulse and the pulse suffers little distortion.

A typical value of pulse duration p is $1\mu\text{s}$. In this case a receiver bandwidth of $\frac{2}{p}$ or 2 Mc/s will pass the $1\mu\text{s}$ pulse with little distortion. In precision radars it may be necessary to reduce distortion of the pulse even more to provide accurate ranging. This means that more of the side frequencies have to be passed by the receiver, and its bandwidth must increase accordingly. Values of bandwidth as great as $\frac{10}{p}$ have been used. For a $1\mu\text{s}$ pulse this would mean a bandwidth of 10 Mc/s. We shall see in the next chapter how such wide bandwidths are obtained.

Bandwidth and Noise

Although we require a *wide* receiver bandwidth for minimum distortion of the pulse this clashes with the requirement of *narrow* bandwidth for minimum noise. Since thermal noise power $= kTB$, where B is bandwidth, it is often necessary to *reduce* the receiver bandwidth to limit noise. We know that thermal noise is a 'white noise' and is uniform throughout the radio frequency spectrum. Thus a 2 Mc/s band of frequencies anywhere within the radio bands produces the same amount of thermal noise. Because of the dependence on bandwidth, a 3 Mc/s band produces more noise.

If the bandwidth of the receiver is *wide* compared with that occupied by most of the signal energy (i.e. greater than about $\frac{2}{p}$) extraneous noise is introduced by the excess bandwidth and the output signal-to-noise ratio decreases (Fig. 12a). On the other hand, if the receiver bandwidth is *narrower* than the main bandwidth occupied by the signal (i.e. less than about $\frac{1}{p}$) the noise energy is reduced but so also is a large part of the signal energy. Again, the output signal-to-noise ratio is reduced (Fig. 12b). There is therefore an *optimum* bandwidth at which the signal-to-noise ratio is maximum. A receiver whose bandwidth is adjusted to produce this maximum signal-to-noise ratio operates as a *matched filter*.

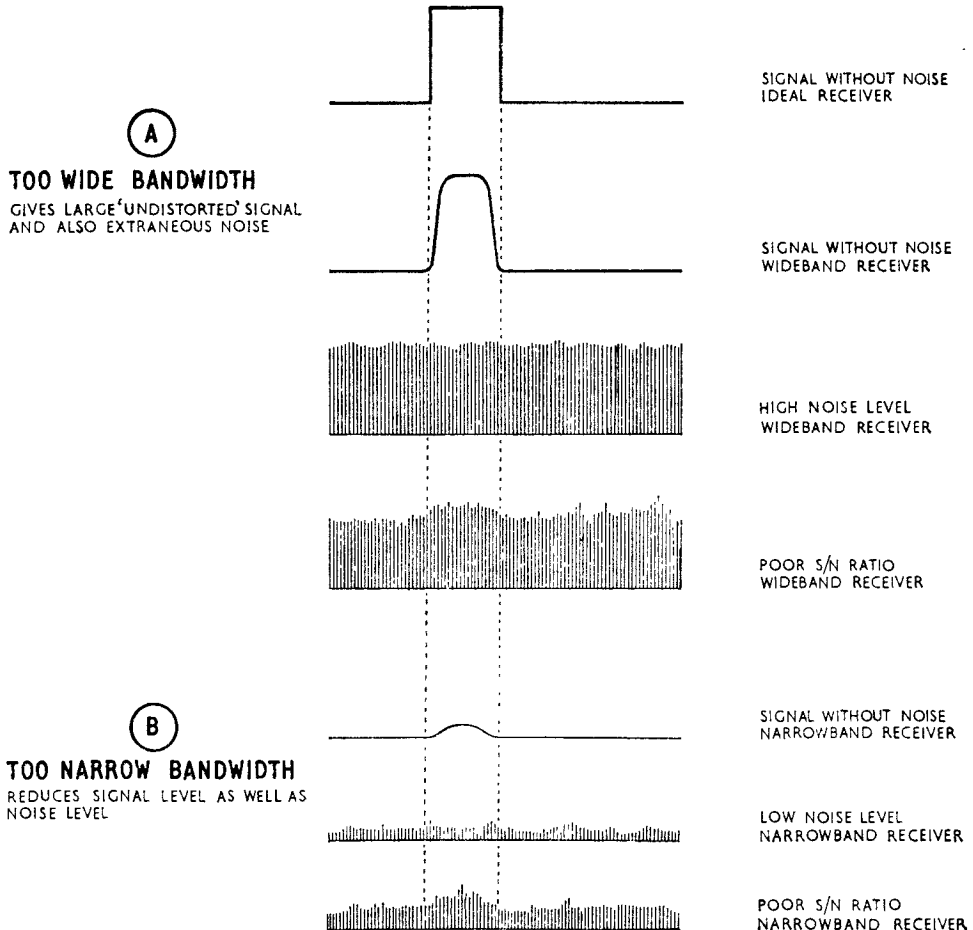


FIG. 12. NEED TO MATCH BANDWIDTH TO BOTH SIGNAL AND NOISE

A precision radar is concerned only with targets at relatively short range. The echo signals are therefore normally large and so also is the input signal-to-noise ratio. Thus the signal-to-noise ratio is not as critical as for long-range radars where the echo is very weak. We can therefore afford to depart from the 'matched filter' idea and increase the bandwidth towards the value $\frac{10}{p}$ mentioned earlier. This brings in more noise, but the signal-to-noise ratio is usually high enough to cope with the increase in noise. At the same time, by passing more of the signal energy we have negligible distortion of the pulse, and range accuracy is thereby improved.

Summary

In this chapter we have discussed the main requirements of a radar receiver. We have shown that a high gain is essential. Gains in excess of 100 db are common in radar receivers so that very weak echoes may be detected. However, the amount of gain that can usefully be applied is limited by the available input signal-to-noise ratio. A high gain receiver is of no value if the signal input is already overwhelmed by noise. The receiver must also be designed with a low noise factor otherwise a reasonable signal-to-noise ratio at the input will be so degraded by the receiver as to be useless at the output. To achieve a low noise factor the first stage in the receiver should have a high gain and a low noise factor. Finally, the bandwidth of the receiver should be adjusted to bring in as much of the signal energy as possible, bearing in mind that this will also bring in more thermal noise. In the next chapter we shall see how these characteristics are obtained.

CHAPTER 2

STAGES IN A RADAR RECEIVER

Introduction

In the previous chapter we discussed the factors which limit the performance of a radar receiver and also the characteristics which are necessary to produce optimum results. In this chapter we shall be looking more closely at the stages in a radar superhet to see how the required characteristics are obtained. The block diagram of the basic radar superhet is redrawn for convenience in Fig. 1.

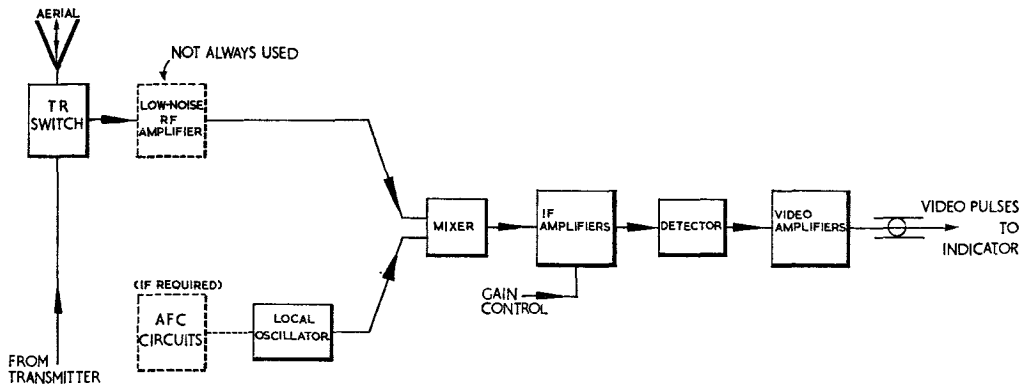


FIG 1. BLOCK DIAGRAM OF BASIC RADAR RECEIVER

Most of the detailed information on the theory of operation of aerials, TR switches, crystal mixers and klystron local oscillators, is found in earlier chapters of this book. Mention is made of them in this chapter merely to provide a logical 'tie-up' of the receiver as a whole. Hence, the treatment of such items in this chapter will be in the form of a summary, with references back to appropriate pages as necessary.

Aerial

The type of aerial used depends to a large extent upon the job that the radar has to do and also upon the frequency at which it is operating. At u.h.f. and lower frequencies an *array* type of aerial is common (p. 237). At microwave frequencies a *reflector* type of aerial may be used (p. 317). In most primary radars the aerial is *common* to both transmitter and receiver.

TR Switch

A TR switching system is necessary to ensure that the aerial is connected to the transmitter for the duration of the pulse and to the receiver for the interval between pulses. The TR switch must also prevent the high power from the transmitter entering the receiver and causing damage. The basic principles of TR switching are dealt with on p. 308. The following notes provide a summary of one basic form of TR switching.

Fig. 2 shows a simple system of branch-arm TR switching using shunt-mounted TB and TR cells. Although open-wire transmission line is shown the system also applies to waveguide and co-axial cable.

During transmission, both cells ionize and the resulting short-circuits at A and B are transformed by $\frac{\lambda}{4}$ lengths of line to effective open-circuits *across* the line at points C and D respectively.

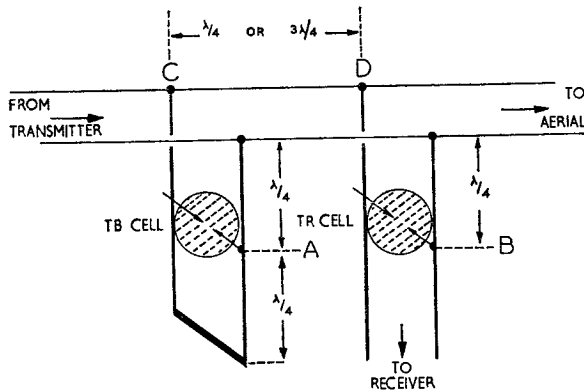


FIG 2. SIMPLE BRANCH-ARM TR SWITCHING

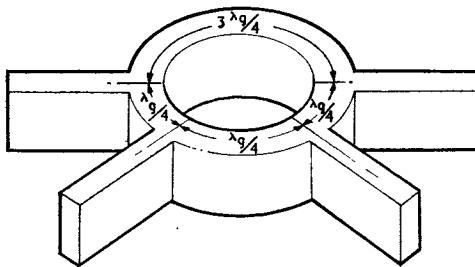
'sees' a very high impedance towards the transmitter from D. Thus, received signals from the aerial pass down the low impedance branch towards the receiver but very little received power passes point D towards the transmitter.

Other devices which are used in TR switching systems, usually in conjunction with TR, TB and pre-TR cells, are illustrated in Fig. 3. They are:

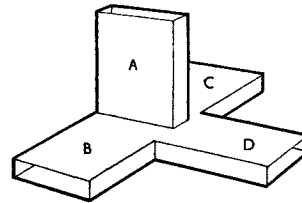
Hence negligible power enters the shunt-arms—only enough to keep the cells ionized. The receiver is therefore protected and the vast majority of the transmitter power passes to the aerial for radiation.

When the transmitter pulse ends, the cells quickly de-ionize. We then have open-circuits at A and B. The short-circuit at the end of the TB arm reflects over half a wavelength as a very low impedance across C. The low impedance at C is further transformed

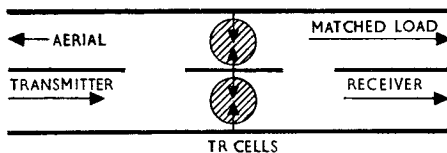
by a $\frac{\lambda}{4}$ length of line so that the signal



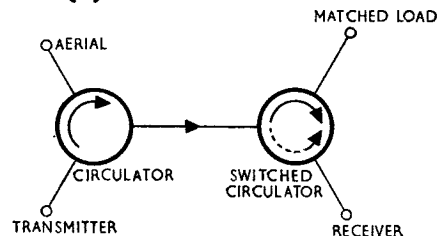
(A) HYBRID RING



(B) MAGIC-T



(C) SHORT-SLOT HYBRID COUPLER



(D) FERRITE SWITCH

FIG 3. TR SWITCHING DEVICES

- a. Hybrid ring, or rat-race (p. 300).
- b. Hybrid tee, or magic-T (p. 299).
- c. Short-slot hybrid coupler (p. 313).
- d. Ferrite switch (p. 315).

RF Amplifier

We have already noted that not all radar superhets use an r.f. amplifier stage. They are common in low-frequency radars but until recently they were seldom found in microwave receivers. The main reason for this is that amplification at very high frequencies is difficult to obtain with conventional circuits. Even where it could be obtained the noise factor tended to be excessive. It is necessary for an r.f. amplifier to have a low noise factor and a good gain. We shall see that these characteristics can readily be obtained at the lower radar frequencies by good circuit design. Only in the last few years, however, have they been obtained at microwave frequencies by the use of such devices as the parametric amplifier, travelling-wave tube, maser and tunnel diode. It can therefore be expected that future radars will employ one or the other of these devices as r.f. amplifiers.

Because a triode does not suffer from partition noise, it is much less noisy than a pentode. At the lower radar frequencies (up to about 1,000 Mc/s) circuits can therefore be designed around a triode to produce a low-noise r.f. amplifier. A basic circuit is illustrated in Fig. 4. The signal is matched to the receiver input by a tap on the coil of the tuned circuit L_1C_1 . The signal is then developed between grid and cathode of the valve and the amplified output is developed across the tuned anode load L_2C_2 . To counteract Miller feedback through the interelectrode capacitance C_{ag} a neutralizing circuit L_3C_3 is inserted.

Another circuit which is much used as a low-noise r.f. amplifier at frequencies up to about 1,000 Mc/s is the *cascode circuit*, shown in Fig. 5. This circuit uses two triodes in *cascode* giving similar gain to that of one pentode but with a much lower noise factor. The first stage V_1 is a low-noise neutralized grounded-cathode circuit similar to that of Fig. 4. The second stage is a

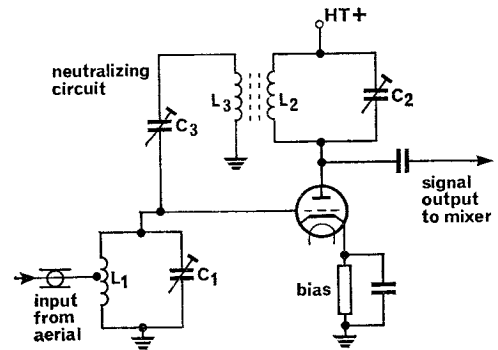


FIG 4. NEUTRALIZED TRIODE AS LOW-NOISE RF AMPLIFIER

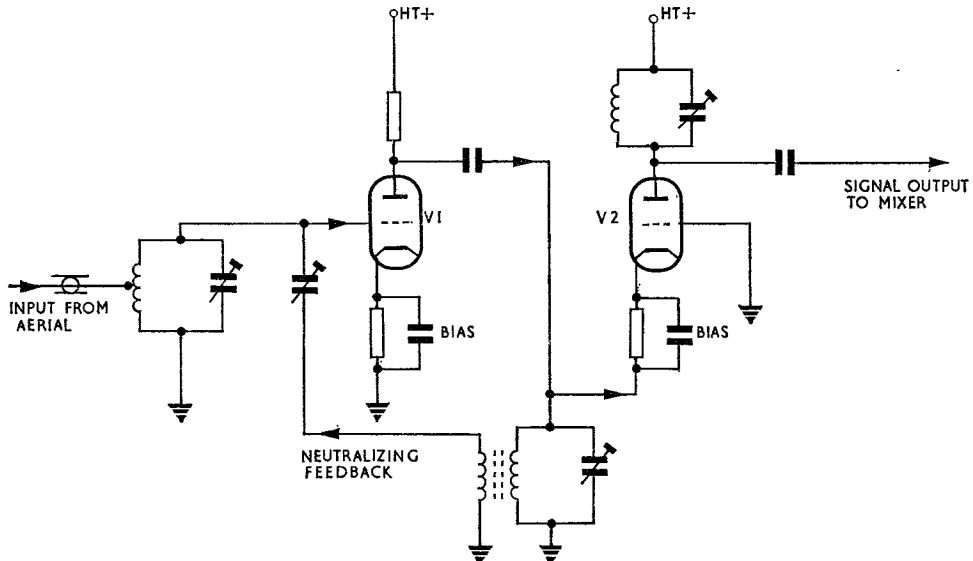


FIG 5. CASCODE CIRCUIT AS LOW-NOISE RF AMPLIFIER

grounded-grid amplifier (see Part 1B p. 334). The grounded-grid amplifier has a gain slightly greater than that of a grounded-cathode amplifier and it is very stable due to the earthed grid between output and input. It is seldom used on its own, however, because it has a low input impedance and this tends to damp the aerial input. A correspondingly large signal power is then required to develop a given voltage across the input terminals. When the two stages are used in cascade this disadvantage is overcome, since a grounded-cathode circuit has a *high* input impedance. In practice V_1 is operated as a low gain amplifier to ensure stability and correct matching. V_2 has a much higher gain, and the two stages together have a gain approaching that of a pentode but without the noise of a pentode.

The valves used in the circuits of Fig. 4 and Fig. 5 are very often disc-seal (planar) triodes. The corresponding tuned circuits are concentric lines. An example of a grounded-grid r.f. amplifier using such components is shown in Fig. 6.

In microwave receivers the high gain and low noise factor needed in the first stage of a receiver can now be obtained by using:

- Parametric amplifiers (p. 348).
- Travelling-wave tubes (p. 263).
- Masers (p. 355).
- Tunnel diodes (p. 342).

A typical input circuit of a microwave receiver using a travelling-wave tube as an r.f. amplifier is shown in Fig. 7. The incoming signal from the aerial passes through the de-ionized TR cell and is applied to the helix of the travelling-wave tube. The amplified output, from the other end of the helix, is then applied through a tuned cavity to the mixer. A travelling-wave tube is a broadband device, and since noise is proportional to bandwidth, the noise factor tends to be rather high (about 7 db). This is reduced by the tuned resonant cavity which passes only the required band of frequencies so that extraneous noise is eliminated. A noise factor of about 3 db results. The gain of the travelling-wave tube amplifier is high (about 25 db) so the two requirements of low noise factor and high gain have been met.

Mixer

The purpose of the mixer is to combine the local oscillator and signal frequencies in such a way as to produce an output at their *difference* frequency (arranged to be the i.f.). Additive mixing

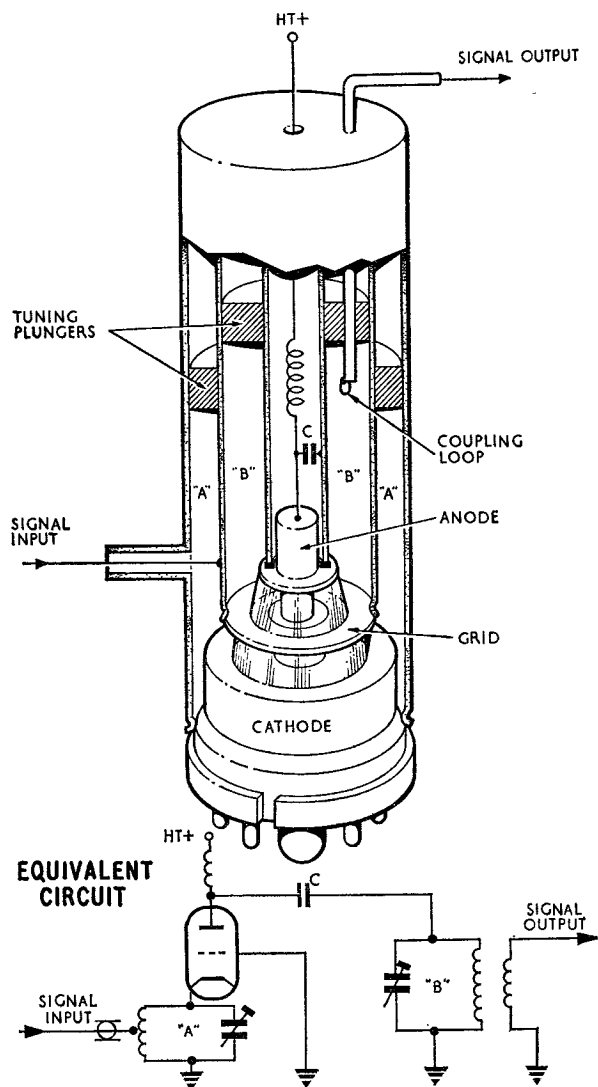


FIG 6. GROUND-GRID RF AMPLIFIER USING DISC-SEAL-TRIODE

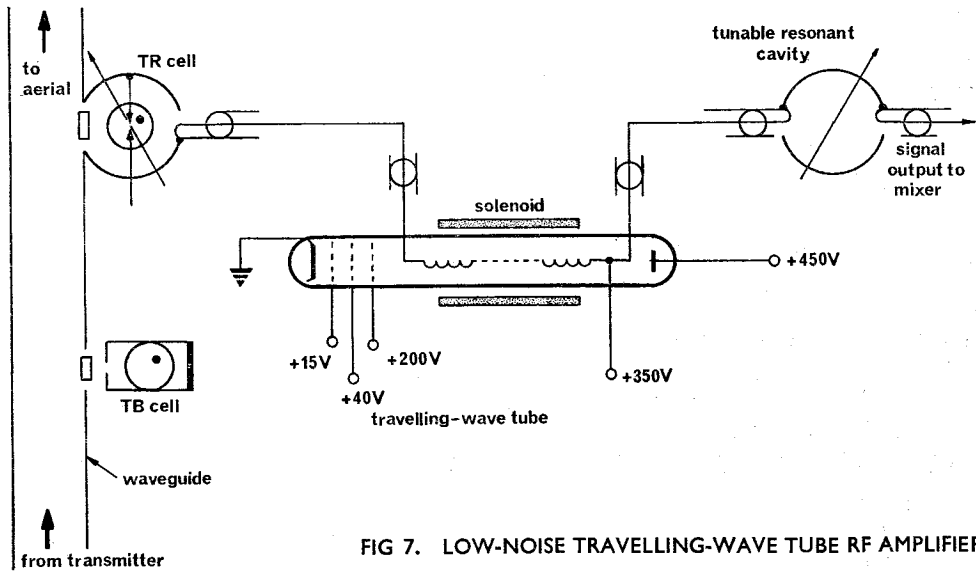


FIG 7. LOW-NOISE TRAVELLING-WAVE TUBE RF AMPLIFIER

is used because the multigrid valves necessary for multiplicative mixing introduce excessive (partition) noise.

At the lower radar frequencies a conventional triode anode-bend detector may be used as the mixer. A large cathode bias resistor is used to bias the valve to the bottom bend of its $I_a V_g$ characteristic. Under these operating conditions the valve functions as a square law detector (Fig. 8a).

A typical circuit is shown in Fig. 8b. The amplified signal from the r.f. amplifier is applied to the mixer grid via a 15pF capacitor. The local oscillator input is similarly coupled but by a much smaller (1pF) capacitor. The choice of capacitor values reduces undesirable coupling between signal and local oscillator circuits and also ensures that the local oscillator input does not swamp the signal. The correct bias is obtained by using a large 3.3k Ω cathode bias resistor. The two inputs beat together at the mixer grid and produce a component at their difference frequency (the i.f.) in the anode current. This i.f. component is selected by the anode tuned circuit (suitably damped by the 1k Ω resistor to provide the required bandwidth) and applied via the 100pF coupling capacitor to the i.f. amplifier stages.

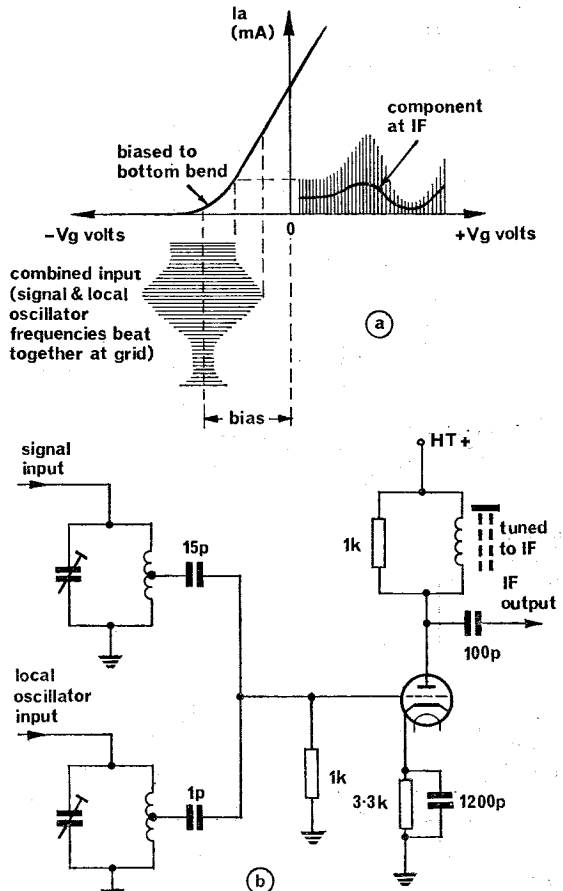


FIG 8. TYPICAL MIXER CIRCUIT USED AT THE LOWER RADAR FREQUENCIES

Very often the valve and associated tuned circuits are of the disc-seal, concentric line type. The arrangement is then similar to that shown in Fig. 6.

At microwave frequencies, the usual problems associated with valves (transit time effects, lead inductance, interelectrode capacitances, induced grid noise) virtually prohibit the use of valves as mixers. At such frequencies a semiconductor crystal diode is used in a diode detector circuit. The crystal mixer is considered in detail on p. 305. In some equipments a single crystal mixer is used; in others, two crystals in a balanced mixer circuit are used to reduce local oscillator noise.

Fig. 9a illustrates a typical schematic diagram of a microwave receiver using a single crystal mixer as the first stage. The incoming signal from the aerial passes through the de-ionized TR cell to the crystal mixer where it combines with the output from the klystron local oscillator. The resultant output from the mixer contains a component at the i.f. and this is applied via coaxial cable to the i.f. amplifiers. A tunable resonant iris in the waveguide bridge controls the amount of local oscillator output reaching the crystal. The equivalent circuit of the crystal mixer is sketched in Fig. 9b.

Very often the *d.c.* component of the mixer output is filtered off separately in the i.f. amplifier stage and is used as a measure of the crystal current. The correct value of crystal current for a given equipment is given in the appropriate Air Publication of that equipment and this value is obtained by adjusting the coupling between the local oscillator and the mixer (by the resonant iris in Fig. 9a).

Local Oscillator

The local oscillator in a radar receiver performs the same function as in any other superhet, i.e. it produces an output which differs from the signal frequency by the i.f. The signal and local oscillator frequencies then combine at the mixer to produce the i.f. Because of the high frequencies used in radar the local oscillator frequency is usually *below* that of the signal by the i.f. As with all generalizations, of course, there are exceptions.

At the lower radar frequencies, the local oscillator is very often a triode arranged in a Colpitts circuit as shown in Fig. 10. The Colpitts circuit has the advantage that it effectively absorbs all the valve interelectrode capacitances into the tuned circuit where their effect on the tuning of the circuit can be accounted for. For this reason almost all triode oscillators operating at very high frequencies use a form of Colpitts. The valve and tuned circuit may be a disc-seal triode, concentric line combination similar to that shown in Fig. 6.

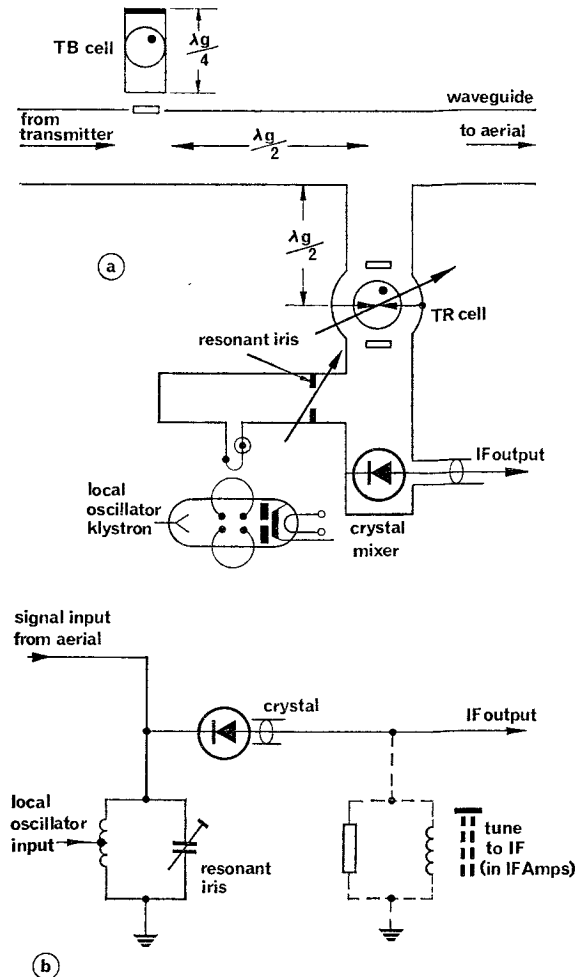


FIG 9. MICROWAVE SINGLE CRYSTAL MIXER

Frequency stability is important in the local oscillator and the use of temperature-compensating components and stabilized power supplies helps to achieve this. In an attempt to obtain good frequency stability, some equipments use a low-frequency crystal-controlled oscillator, followed by several stages of frequency multiplication.

At microwave frequencies the usual local oscillator is a reflex klystron (p. 258). A sketch of a 10 cm reflex klystron, together with its circuit symbol, is shown in Fig. 11a. A typical klystron local oscillator circuit is shown in Fig. 11b.

The frequency of oscillation of the klystron depends upon two factors: the frequency to which its resonant cavity is tuned, and the voltage on its reflector. We can usually vary the

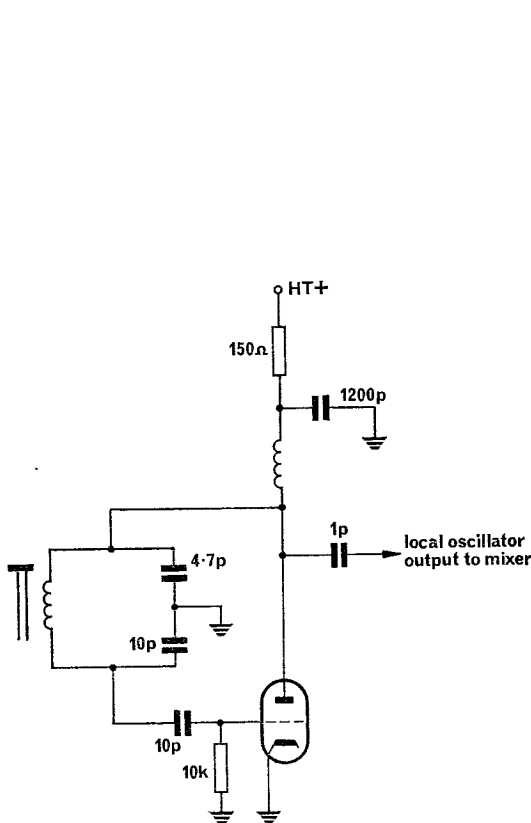


FIG 10. TYPICAL LOCAL OSCILLATOR CIRCUIT FOR THE LOWER RADAR FREQUENCIES

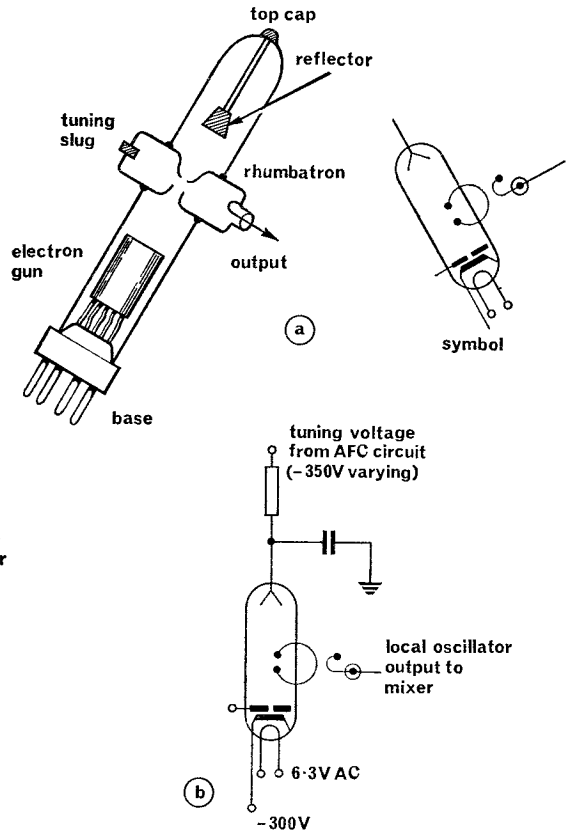


FIG 11. KLYSTRON LOCAL OSCILLATOR

frequency by *mechanical* means over a fairly wide frequency range (often as great as 500 Mc/s). The mechanical tuning is by a variation in the size of the resonant cavity, either by an inductive tuning slug or by distorting the cavity walls.

The klystron is mechanically preset to a frequency which is the i.f. below the signal frequency. Once the frequency has been preset in this way we can then vary the frequency as required by adjusting the voltage of the reflector. This 'electrical' variation of the klystron frequency (typically ± 30 Mc/s) is obtained automatically by using the output of the a.f.c. circuit to vary the reflector voltage. We have already noted that any variation in the frequency of the transmitter magnetron output (and hence the signal input) must be followed exactly by the klystron local oscillator to keep the output from the mixer at the correct i.f. The a.f.c. circuit ensures this.

Automatic Frequency Control (AFC)

The need for an a.f.c. circuit in radar receivers using a klystron local oscillator has already been mentioned. Let us now see how the a.f.c. circuit plays its part in ensuring that the output from the signal mixer is at the correct i.f. A simple block diagram of a basic a.f.c. circuit is shown in Fig. 12.

When the transmitter fires, a very small portion of the transmitted pulse is fed, via an attenuation section of waveguide, to the a.f.c. mixer. Also applied to this mixer we have the output

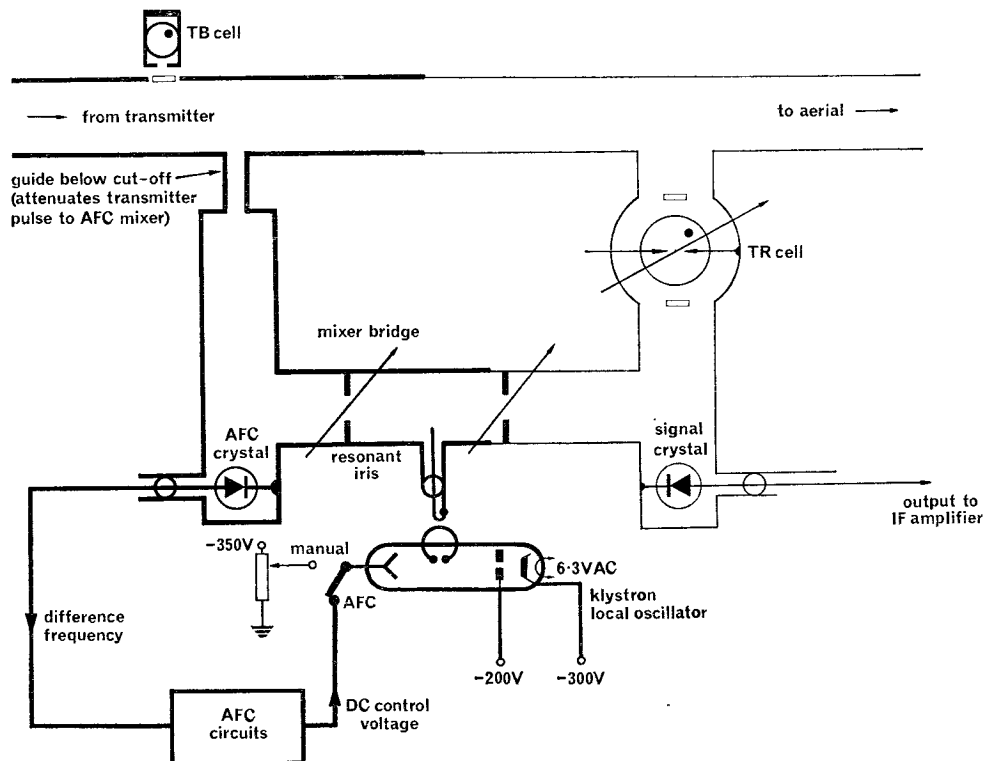


FIG 12. OUTLINE OF AFC SYSTEM

from the klystron local oscillator. The frequency output of the a.f.c. mixer is thus the *difference* between the sample transmitted pulse and the local oscillator, and this difference frequency is fed to the a.f.c. circuits. The output from the a.f.c. circuits is a *negative* d.c. voltage which is applied to the reflector of the klystron when the switch is at AFC. This voltage determines the frequency at which the klystron oscillates. Note that, by means of the switch, the klystron can be disconnected from the a.f.c. circuit and tuned *manually* by adjustments of a potentiometer connected to a negative voltage supply.

If the difference between the transmitted frequency (and hence the signal echo frequency) and the local oscillator frequency is equal to the correct i.f. the resultant output from the a.f.c. circuits is such that the reflector voltage maintains the klystron frequency at its present value. If the transmitter magnetron frequency *increases*, the difference frequency increases (local oscillator *below* signal in frequency) and this causes the output from the a.f.c. circuits to become *more negative*. As we saw earlier in these notes, when the klystron reflector voltage is made more negative its output frequency *increases* (Fig. 13). Similarly, if the magnetron frequency *decreases*, the output from the a.f.c. circuits becomes *less negative* and the klystron frequency also *decreases*.

Thus for any variation in the magnetron frequency the a.f.c. circuits produce a voltage variation which, when applied to the reflector, causes the klystron to shift in frequency by approximately the *same amount* and in the *same direction* as the magnetron. Hence, the difference between the signal frequency and the local oscillator frequency is always very nearly equal to the correct i.f. Because the local oscillator also feeds the main signal mixer, the receiver i.f. is also approximately constant for any variation of transmitter frequency.

In some equipments the difference between the magnetron and klystron frequencies when first switched on produces an i.f. which is outside the bandwidth of the a.f.c. circuits. Hence the a.f.c. circuits cannot produce the required control voltage. To overcome this, a 'sweep oscillator' is usually employed to vary the klystron reflector voltage backwards and forwards until the i.f. produced lies within the a.f.c. circuits bandwidth. The d.c. control voltage from the a.f.c. circuits then stops the action of the sweep oscillator and the circuit operates as described above. In effect, the sweep oscillator causes the klystron to 'lock-on' to the correct frequency and the a.f.c. circuits then take over.

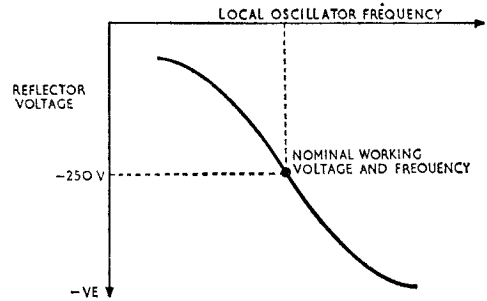


FIG 13. EFFECT OF VARYING KLYSTRON REFLECTOR VOLTAGE

AFC Circuits

Now that we have seen what the a.f.c. system does let us have a look at the circuits which produce the control voltages. The block diagram of the circuit necessary to produce the control voltages is shown in Fig. 14a and a simplified circuit arrangement in Fig. 14b.

The beat frequency from the a.f.c. crystal mixer is applied to V_1 which operates as a conventional wideband i.f. amplifier whose circuits are tuned to the receiver i.f. In practice, two or more such stages are used to provide high gain. The output from V_1 is then applied to the

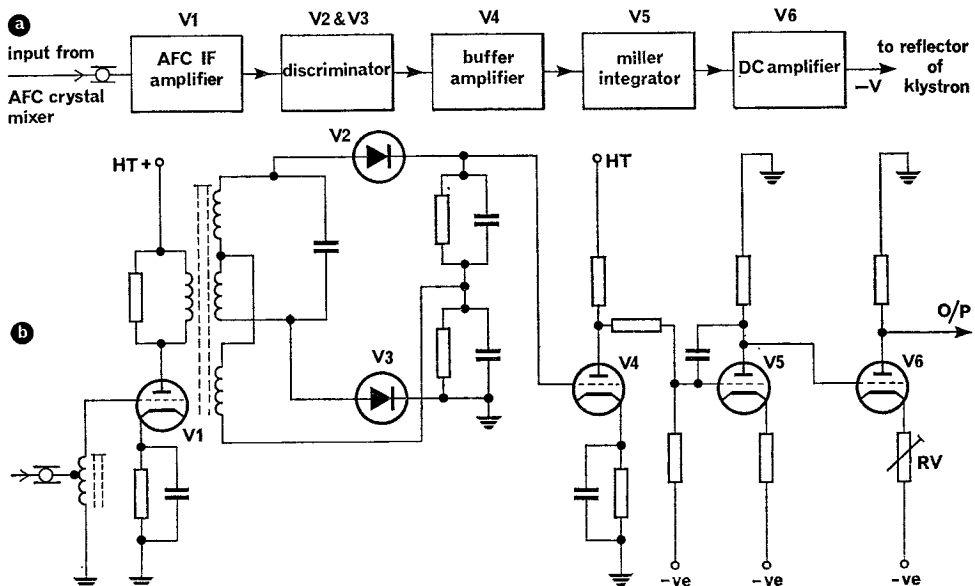


FIG 14. BASIC AFC CIRCUIT

discriminator, consisting of V_2 and V_3 . Many different types of discriminators are used in practice. Some are based on either the Travis or Foster-Seeley circuits which we discussed in Part 1B (p. 426). In Fig. 14b a Foster-Seeley discriminator is shown. When the beat frequency from the a.f.c. mixer is at the correct i.f. the discriminator output is zero. When the input is below the i.f. in frequency the discriminator output is a series of negative pulses at the transmitter p.r.f.; and when the input is above the i.f. the discriminator output consists of positive pulses. After amplification in the d.c. coupled stage V_4 , the discriminator output is fed to the grid of the miller integrator stage V_5 . This stage is necessary to provide the smoothly varying d.c. voltage required to control the frequency of the reflex klystron local oscillator at its reflector. As this voltage must always be negative with respect to the klystron cathode, any stages of d.c. amplification which follow the integrator must be connected so that their output is always negative. In practical circuits more than one stage (V_6) of amplification may be used to provide high gain and thus improve the accuracy of the a.f.c. loop.

When an equipment is first switched on, the i.f. produced may be outside the passband of the simple a.f.c. loop shown in Fig. 14. A more complicated circuit having both 'search' and 'lock' phases is therefore normally used. In the search phase separate aiming voltages are often switched to the grid of the miller integrator V_5 through a relay system controlled by the buffer amplifier V_4 . This causes the klystron reflector to be swept as shown in Fig. 15 until an output is obtained from the discriminator; causing V_4 to switch the relays to the lock condition and the reflector voltage to be controlled by the a.f.c. loop output. In some equipments, the search phase is achieved by mechanical tuning of the klystron with the tuning motor relay controlled.

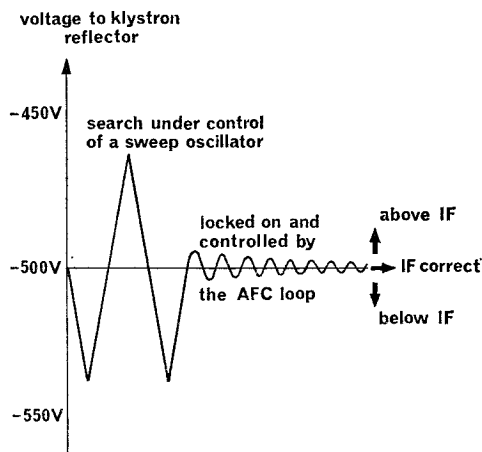


FIG 15. OUTPUT TO REFLECTOR OF KLYSTRON

IF Amplifiers

We have already noted the required characteristics of the i.f. amplifier stages in a radar receiver:

- They must provide a high gain (a gain of 100 db is common).
- They must have a bandwidth necessary to preserve the pulse shape whilst maintaining the required signal-to-noise ratio (a bandwidth of several megacycles per second may be necessary).
- Any tendency to instability (common in high frequency, high gain amplifiers) must be prevented.

In general, for any given amplifier, the product of gain and bandwidth is a *constant*. If we increase the bandwidth, the gain goes down, and *vice versa*. Thus to provide a high gain with a wide bandwidth, many i.f. stages are needed. Six stages are about the minimum that have been used in practice, and some equipments have as many as twelve.

To provide the necessary bandwidth we can place damping resistors across the i.f. tuned circuits; we can use band-pass coupled circuits with greater-than-critical coupling to give a double-humped response; or we can stagger-tune individual stages. In most equipments a combination of these methods is used.

A damping resistance placed across a tuned circuit flattens the response curve and increases the half-power bandwidth. The effect of greater-than-critical coupling between two tuned circuits is considered in Part 1B (p. 137); again the half-power bandwidth is increased relative to that for loose coupling. An example of stagger-tuning is shown by the block diagram of Fig. 16. The i.f. is assumed to be 45 Mc/s. V_1 is tuned to 45 Mc/s, V_2 to 43 Mc/s, V_3 to 47 Mc/s, V_4 to 43 Mc/s, V_5 to 47 Mc/s, V_6 to 45 Mc/s. Damping resistors may also be used, and a very wide bandwidth results (in this example, about 6 Mc/s).

Instability is prevented by adequate decoupling of stages and screening of components, and by careful design and layout. A small amount of negative feedback, usually from undecoupled cathode resistors, may also be used.

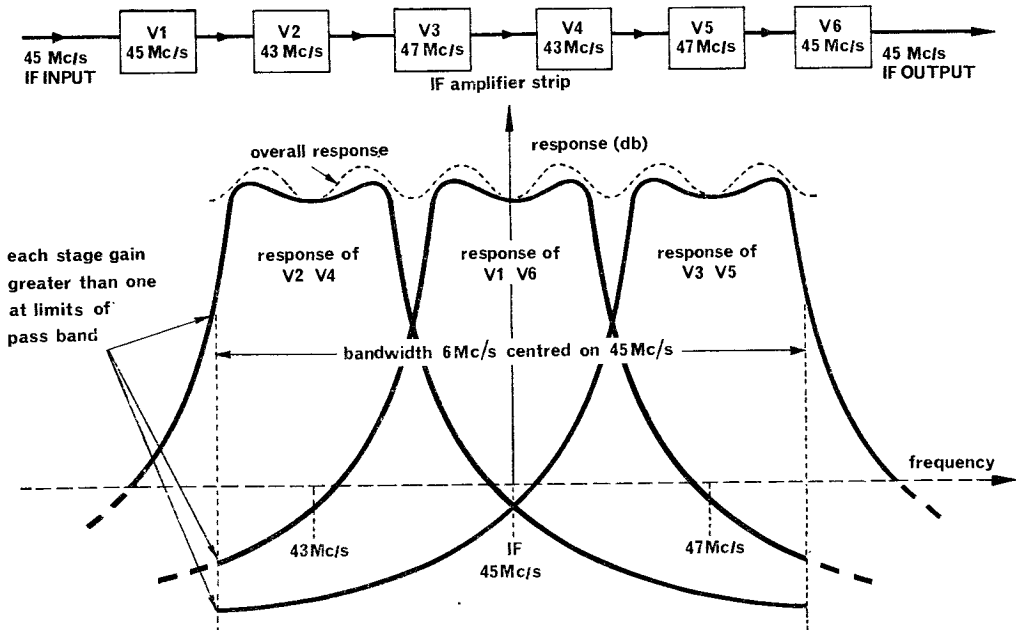


FIG 16. STAGGER-TUNING

Control of gain is usually achieved by varying the control grid bias of the first two or three i.f. amplifiers. This may be done manually by applying a variable negative bias to the control grids from a potentiometer, through the normal decoupling networks. AGC, if used, can be applied in the same way.

Very often, to improve the receiver noise factor, the first one or two i.f. amplifier stages use low-noise triodes in a neutralized or cascode circuit. These circuits have already been considered in Figs. 4 and 5. This practice is particularly important in receivers in which the crystal mixer is the first stage. In such cases the low-noise i.f. amplifiers are usually situated in the r.f. head of the equipment.

IF amplifier circuits differ considerably from one equipment to another. Fig. 17 shows two typical circuits, both assumed to be operating at an i.f. of 45 Mc/s. At this frequency the self-capacitance of the windings, plus stray capacitance, is sufficient to resonate with the inductance to form the tuned circuits. Tuning during re-alignment, is by variable cores in the inductors. Both circuits use stagger-tuning.

The circuit in Fig. 17a illustrates single-tuned coupling. The $1.5k\Omega$ anode load resistor R_1 of the first stage is effectively in parallel with the tuned circuit inductance L_2 of the second stage

and this provides damping of the tuned circuit. The other stages are similarly damped to provide the required bandwidth.

The circuit in Fig. 17b is a double-tuned circuit. The required bandwidth is achieved by adjusting the coupling between the two circuits, by the damping resistor across the primary circuit, and by stagger-tuning.

Each anode is adequately decoupled (screen grids also, although not shown) and negative feedback from undecoupled cathode resistors is also applied. The gain control voltage is applied to the grids through the decoupling networks. As the gain is varied, the alignment of the tuned circuits is upset due to the variation of the input capacitance of the stage as a result of Miller effect. This effect is reduced by negative feedback.

Detector

In a broadcast receiver the incoming signal is demodulated to obtain an output identical with the original modulating waveform at the transmitter. Similarly, in a radar receiver, the pulse-modulated i.f. signals must be detected to obtain video pulses.

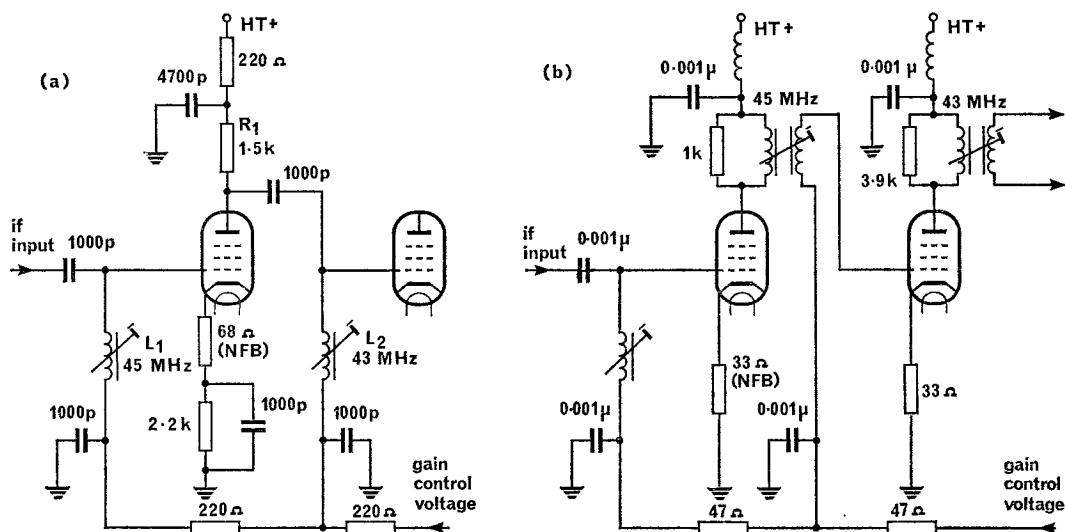


FIG 17. TYPICAL IF AMPLIFIERS, SIMPLIFIED CIRCUITS

Diode detectors, either semiconductor or valve, are used. A basic circuit arrangement is illustrated in Fig. 18a. The diode load components R_1C_1 must have values much smaller than those in a broadcast receiver. Fig. 18b shows that there are two reasons for this:

- C_1 must charge rapidly through the resistance R_d of the conducting diode to maintain the steep leading edge of the pulse waveform.
- The time constant C_1R_1 must be short enough to allow C_1 to discharge quickly through R_1 at the end of the pulse to prevent a long tail.

It is also necessary for the conducting resistance R_d and the self-capacitance C_d of the diode to be small compared with the values of R_1 and C_1 respectively. If R_d is too large the detected voltage developed across R_1 is correspondingly smaller, since R_1 and R_d form a potential divider across the input tuned circuit. If C_d is large its reactance is low at the high i.f. used in radar receivers and this causes C_d to shunt the diode. Small values of C_d and R_d can best be obtained by using small semiconductor germanium junction diodes.

In common with the detector in a broadcast receiver a filter network is necessary in the output of the detector to remove unwanted components. The unwanted components are the i.f. and the d.c. It is more difficult to achieve adequate filtering in a radar receiver because, to preserve the shape of the video pulse, components from a few cycles per second up to several megacycles per second have to be passed. In fact, a band of frequencies equal to *one half* of the i.f. bandwidth must be passed. Because of this, detector filter circuits vary considerably from one equipment to another.

The output may be either positive-going or negative-going depending upon the diode connections. A positive-going output is indicated in Fig. 18 but in practice both positive and negative detectors are common.

The d.c. component of detection is a measure of the output of the receiver. Such measurements are needed in calculating the receiver noise factor (*see* p. 401) and also in receiver alignment. The d.c. output can be measured by applying the output voltage developed across the diode load to a filter circuit, which removes all the components except the d.c.

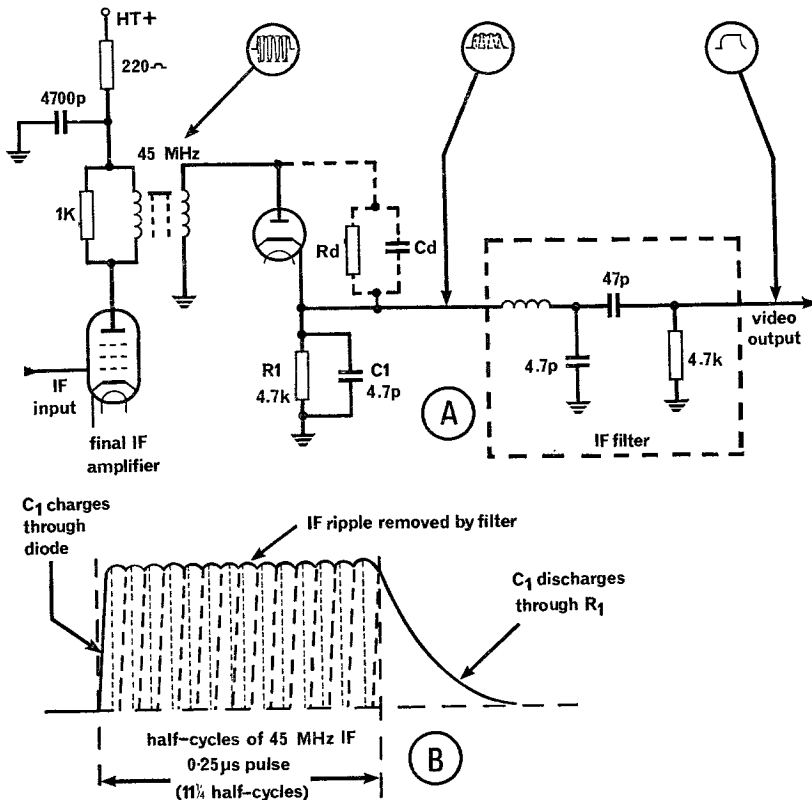


FIG 18. BASIC VIDEO DETECTOR

Video Amplifiers

The purpose of the video amplifier is to amplify the output from the detector to a sufficiently high level for satisfactory operation of the c.r.t. display. In doing so, the video amplifier must introduce as little distortion as possible. The basic theory of video amplifiers is considered in AP 3302 Part 1B p. 319. We shall therefore merely review the subject here and show a typical circuit used in a radar receiver.

The output from the detector is a video pulse. This can be resolved into a series of sine waves covering a very wide frequency range. To avoid distortion of the pulse shape the video amplifier must respond equally to all the sine wave components within the pulse, i.e. the video amplifier must have a *wide bandwidth*. Since one of the sidebands of the received signal has been removed by detection, the bandwidth required by the video amplifier is *half* that of the i.f.

amplifiers. For a 5 Mc/s i.f. bandwidth, all frequencies up to 2.5 Mc/s should be passed equally by the video amplifier.

If we analyse the construction of a pulse from its series of sine waves we shall notice that it is the *low frequency* components which determine the flatness of the top of the pulse and the *high frequency* components which determine the steepness of the leading and trailing edges. Hence if the pulse, in passing through the video amplifier, develops a 'sag', this indicates that the low frequency response of the amplifier is inadequate (Fig. 19a). If the leading or trailing edges develop an exponential rise or fall this indicates that the high frequency response of the amplifier is poor (Fig. 19b).

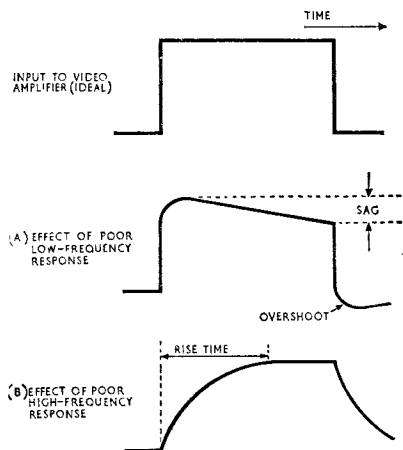


FIG 19. EFFECT ON PULSE OF POOR VIDEO FREQUENCY RESPONSE

across the voltage output of V_1 . The lower the frequency, the higher is the reactance of C_c . Thus more of V_{a1} appears across C_c at the lower frequencies and less across R_g . V_{g2} thus decreases with decrease in frequency and the top of the pulse sags.

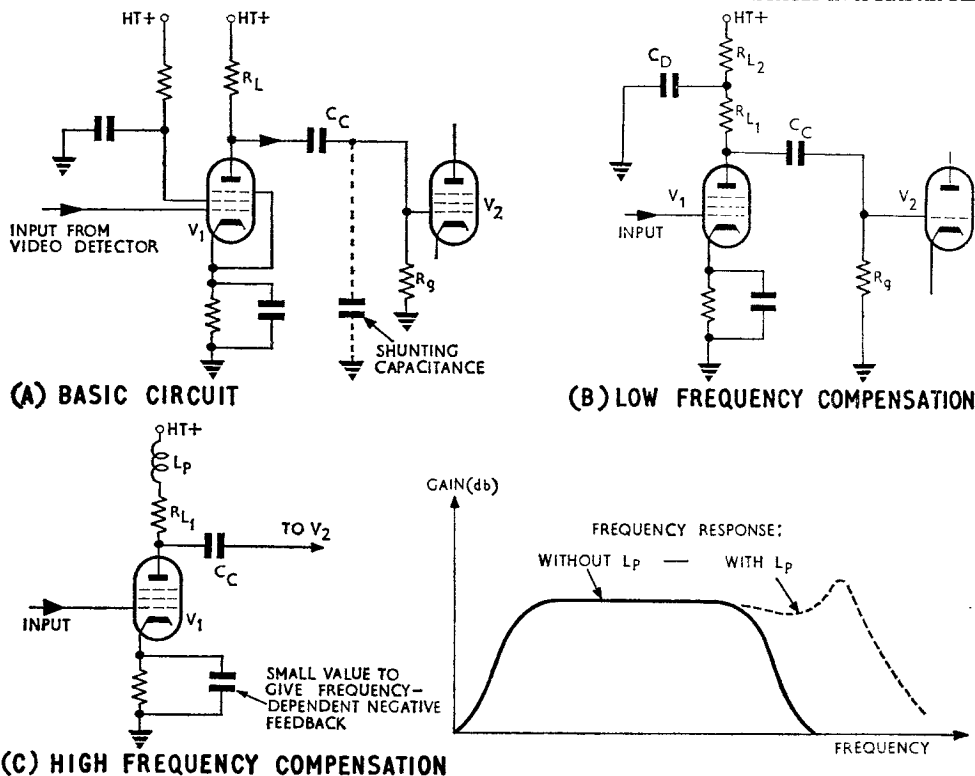
The frequency at which this effect becomes pronounced can be decreased by making C_c large. It can, in fact, be eliminated by using a DC coupled amplifier. However, the usual method of low frequency compensation is shown in Fig. 20b. The anode load of V_1 is split and the top part of the load R_{L2} decoupled by C_D (the same value approximately as C_c). At low frequencies as the reactance of C_c increases so does that of C_D and C_D becomes *less effective* in decoupling R_{L2} . Hence R_{L2} begins to form part of the anode load and the resultant increased gain with decreasing frequency compensates for the loss across C_c .

Poor high-frequency response results from wiring and interelectrode capacitances effectively shunting the output of V_1 . The higher the frequency, the lower is the reactance of the shunting capacitance, and the voltage developed across R_g decreases. This effect reduces the steepness of the leading and trailing edges of the pulse.

The shunt capacitance can be reduced by the use of pentodes (to reduce interelectrode capacitance effects) and by careful design of the amplifier. A reduction in the value of R_{L1} also helps because then the shunt capacitance can charge and discharge more rapidly to preserve the steep leading and trailing edges. Of course we lose something by doing this—a small R_{L1} means a low-gain amplifier.

The usual method of high-frequency compensation is shown in Fig. 20c. A small peaking inductor L_p is inserted in series with the load R_{L1} . The reactance of the coil increases with increase in frequency, thus increasing the effective anode load and hence the stage gain. This tends to compensate for the losses at the higher frequencies. The value of L_p is usually such that it resonates with the shunt capacitance towards the top end of the frequency range and this extends the high-frequency response of the amplifier.

Another method of high frequency compensation which is sometimes used is *frequency-dependent negative feedback*. The cathode resistor may be shunted by a small value capacitor.



At the lower end of the frequency range this capacitor has a high reactance and is ineffective as a decoupling capacitor so that negative feedback from the cathode resistor is operative. At the higher frequencies the capacitor has a lower reactance and tends to decouple the cathode resistor. Negative feedback is thus progressively reduced as the frequency increases, and gain increases with frequency.

In most radar receivers the final video stage is a cathode follower. Although a cathode follower gives no voltage gain, it has a low output impedance which facilitates matching to the coaxial cable carrying the video signal to the indicator.

The simplified circuit of a typical video amplifier is shown in Fig. 21. The first stage operates as a normal video amplifier giving a gain of about ten. High-frequency compensation is obtained by inserting a $130\mu\text{H}$ peaking inductance and also by the use of frequency-dependent negative feedback via the 180Ω cathode resistor. There is no low-frequency compensation but the large value of coupling capacitor ($0.1\mu\text{F}$) ensures satisfactory low frequency response. The second stage operates as a cathode follower to provide the correct source impedance for the c.r.t.

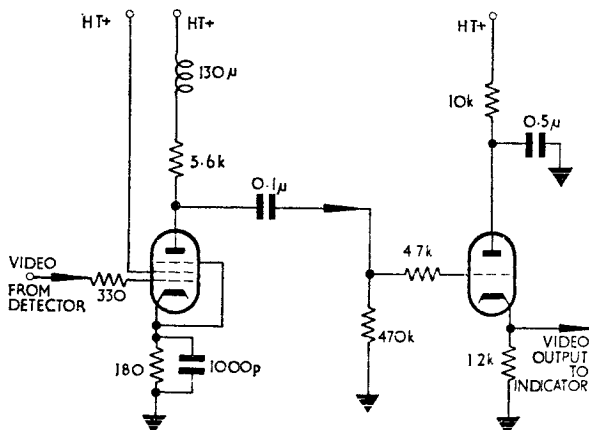


FIG. 21. PRACTICAL VIDEO AMPLIFIER

Receiver Paralysis

Although we take precautions to prevent the high-powered transmitter pulses from entering the receiver circuits there is always a small amount of 'break-through'. Unless we take further precautions each break-through pulse will be amplified by the receiver to such an extent that

coupling and bias capacitors in the final stages of the receiver become charged whilst the transmitter is firing. This saturates or paralyses the receiver so that, even after the transmitter pulse has ended, the receiver cannot handle signals until the capacitors discharge. It therefore takes the receiver some time to settle down to normal conditions after each transmitter pulse and during this time targets near the radar may be missed.

To prevent receiver paralysis, some radars apply a negative-going suppression pulse from the master timing unit to the suppressor grids of the final two or three i.f. amplifiers in the receiver. As shown in Fig. 22, the suppression pulse is coincident in time with the modulating pulse to the transmitter and this cuts off anode current in the suppressed valves for the duration of the output pulse. Saturation of the receiver is thus prevented.

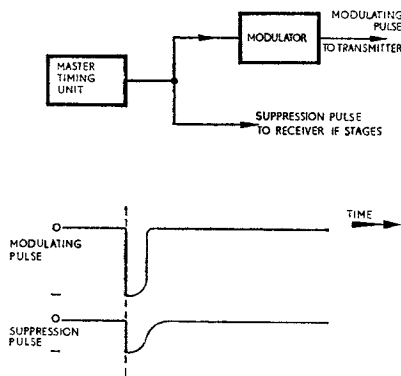


FIG. 22. PREVENTION OF RECEIVER PARALYSIS

Summary

Apart from the crystal mixer and the semiconductor diode detector, all the circuits considered in this chapter have used valves. As we know, the modern tendency is to do away with valves as much as possible and replace them with their semiconductor equivalents. There is no reason why this cannot be done for the whole receiver, and completely 'solid state' radar receivers will become the rule rather than the exception in future radar equipments. We are often going to meet the term 'solid state' in the future. Basically it applies to an equipment or a circuit in which there are no thermionic devices, all of them having been replaced by semiconductor devices. The advantages of this in terms of the size and weight of an equipment, its reliability, power savings, and so on, are very great.

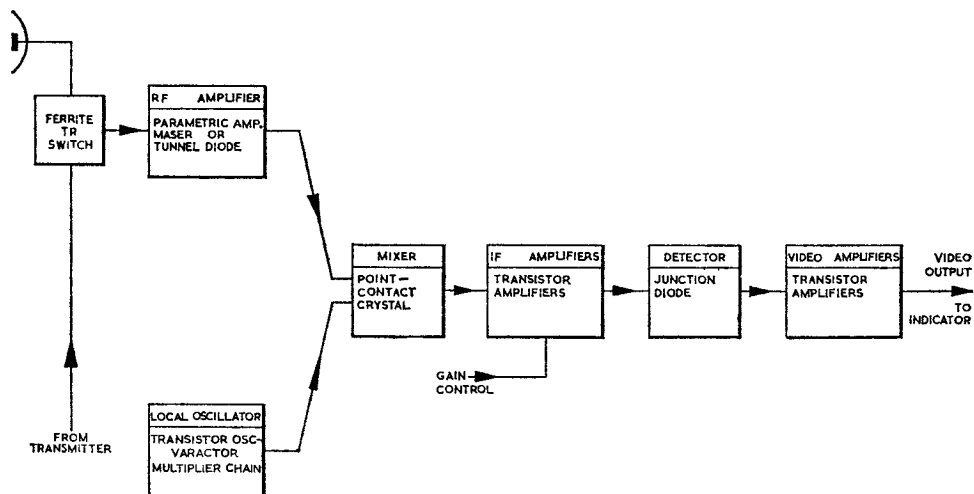
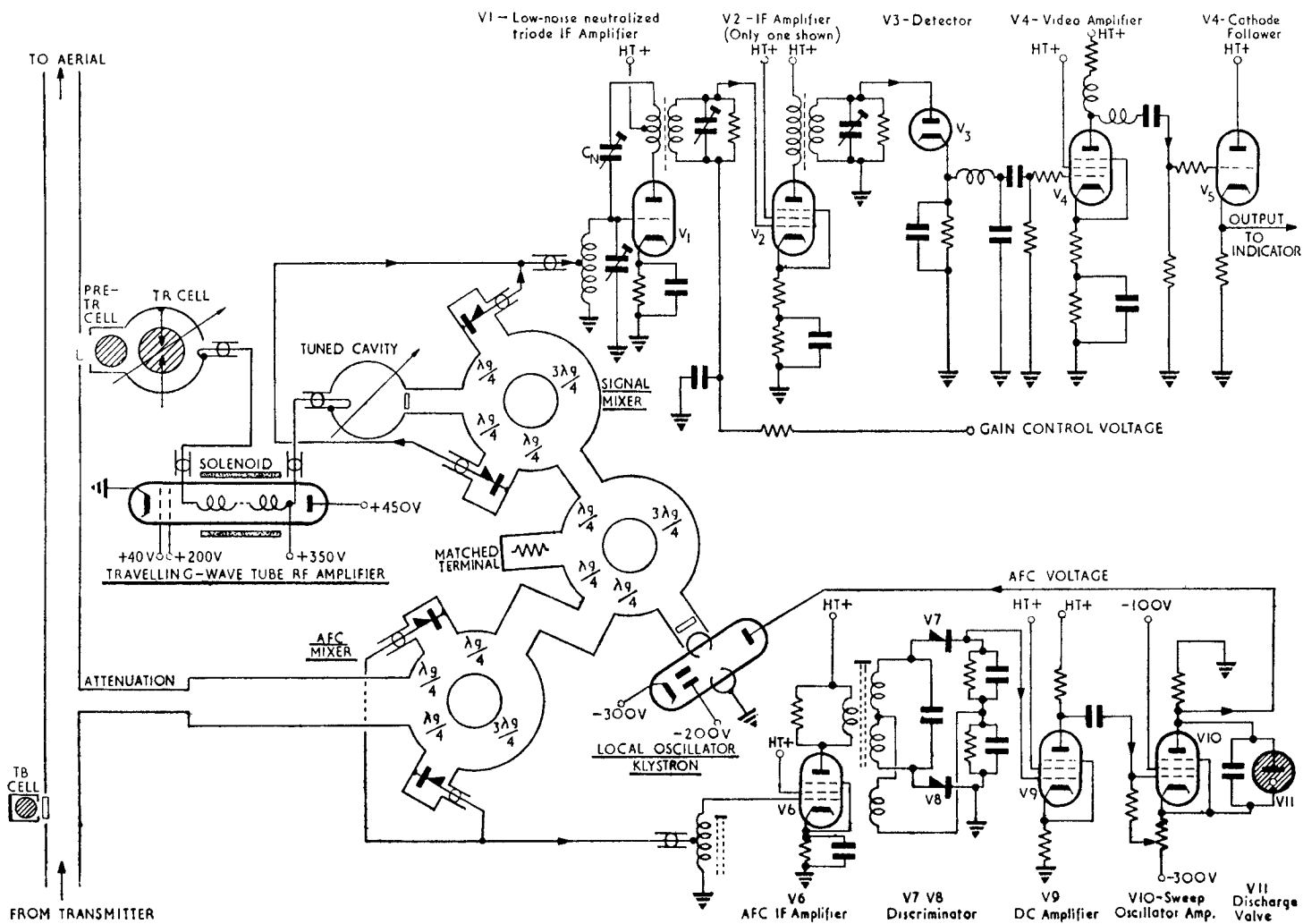


FIG. 23. SOLID STATE RADAR RECEIVER

If we quickly examine the various stages in a radar receiver we shall see that there is no stage which cannot use a solid state device. This is illustrated in Fig. 23. Note however, that, in general, the principles we have considered in this chapter also apply to the solid state receiver.

To conclude this chapter, and to consolidate what we have done, we can bring together all the various stages that we have discussed separately and build up a basic radar receiver. This is done in Figs. 24 and 25. Fig. 24 overleaf illustrates the simplified circuit of a receiver suitable for use at the lower radar frequencies; Fig. 25 shows that of a centimetric radar receiver.





CHAPTER 3

REDUCTION OF CLUTTER

Introduction

So far we have considered the effects of *noise* on receiver performance and have outlined the steps taken to improve the signal-to-noise ratio and the receiver noise factor. However, the required echo signal has also to compete against all forms of *clutter* in addition to noise. We considered clutter very briefly in Section 1. In this chapter we shall examine the subject in more detail and, in particular, discuss the steps that can be taken to reduce the effects of clutter.

Clutter

Clutter may be defined as *confused unwanted echoes* on a radar display. We must, however, be clear on what we mean by 'unwanted echoes'. All objects in the path of a radar beam reflect energy to some extent. Thus we get echoes from the ground, from the sea, from hills, built-up areas, clouds, rain, ships, aircraft, and the like. Something which may be an unwanted echo to one type of radar may be the wanted echo to another type. For an early-warning search radar as used in the RAF, the wanted echoes are those produced by aircraft or missiles; all other objects produce unwanted echoes and represent clutter. For meteorological radar, on the other hand, everything other than clouds or rain produces clutter. For bombing radar, the towns and hills and rivers produce the wanted echoes, i.e. the radar map.

In early-warning search radar, echoes produced by anything other than aircraft or missiles are unwanted and produce clutter. We are then faced with the problem of *removing* the clutter on the c.r.t. display (Fig. 1). Such clutter may be so pronounced, especially from objects near the ground radar, as to completely hide the wanted echoes. In addition, if the clutter is strong enough it can *saturate* the receiver and the time taken for the receiver to recover may mean that some targets have been missed.

Where only echoes from moving targets are of importance, clutter due to reflections from stationary objects can be considerably reduced by using *moving target indication* (m.t.i.) radars. This is discussed in Section 7.

For a non-m.t.i. radar, some reduction of *weather clutter* can be obtained by operating the radar at a low frequency (p. 228). Where this is not possible, *circular polarization* will help in reducing this form of clutter (p. 325).

In this chapter we shall consider modifications and additions that can be made to the circuits of a normal radar receiver in an attempt to reduce all forms of clutter.

It is often difficult to distinguish between noise and clutter. Both produce similar effects on the c.r.t. display. However, clutter is caused by transmissions *from the radar itself*, whereas noise is present at all times. Because of this, the appearance of clutter on the display will remain substantially the same from sweep to sweep, whilst that of noise will vary in a random manner.

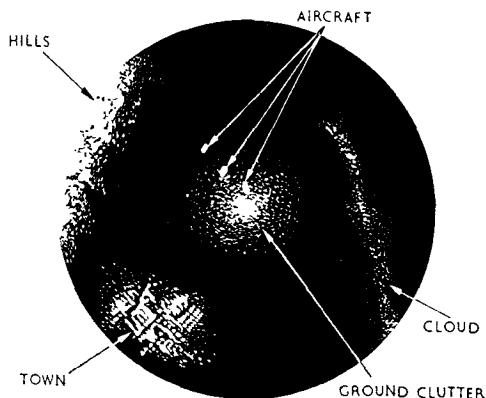


FIG. 1. EFFECT OF CLUTTER ON A PPI DISPLAY

Fast Time Constant (FTC) Circuit

Where clutter is of an 'extended' nature, so that it fills a large area of the display, some reduction in its effect can be obtained by inserting a high-pass filter between the detector and video amplifier stages of the receiver. The filter usually consists of a short CR differentiating circuit which can be switched in or out of the circuit as required and whose time constant is of the same order as the pulse duration of the radar.

A typical arrangement is shown in Fig. 2. Under normal conditions the f.t.c. circuit is switched off and C_1 is short-circuited. R_1 and R_2 in parallel then form the detector load.

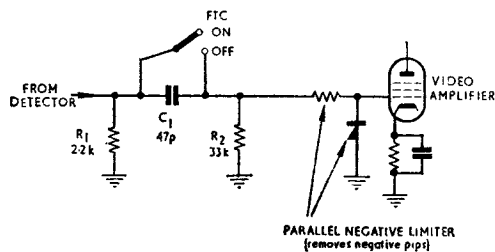


FIG. 2. BASIC FTC CIRCUIT

In conditions of extended clutter the f.t.c. circuit is switched on and C_1 is inserted. The detector load now consists of R_1 only, whilst C_1 and R_2 form a short CR differentiating circuit across the input of the video amplifier (CR time constant is $1.55\mu\text{s}$).

To see the effect of this circuit, let us suppose that we are receiving target echoes from an aircraft under conditions of extended cloud clutter. The p.p.i. presentation with the f.t.c. circuit switched off may then be as shown in Fig. 3a. The cloud produces a large clutter echo which partially obscures the target echo. The waveform of the output from the detector may have the form shown in Fig. 3b. The cloud echo varies in amplitude and is weaker than the aircraft echo, but it is of *relatively long duration*. Thus the detector output consists of a short pulse (due to the aircraft) superimposed on a long pulse (due to the cloud). If now we switch the f.t.c. circuit on, the detector output is differentiated by C_1R_2 to produce the familiar positive and negative pips of voltage (Fig. 3c). Only the positive pips are amplified in the video stages and the resulting p.p.i. presentation with f.t.c. in operation is as shown in Fig. 3e.

Since the f.t.c. circuit removes some of the desired signal energy as well as that due to clutter, the f.t.c. is switched into the receiver only when needed.

It is worth noting that the f.t.c. circuit is successful in reducing all forms of 'long pulse' interference. The interference could take the form of a c.w. signal, either modulated or unmodulated.

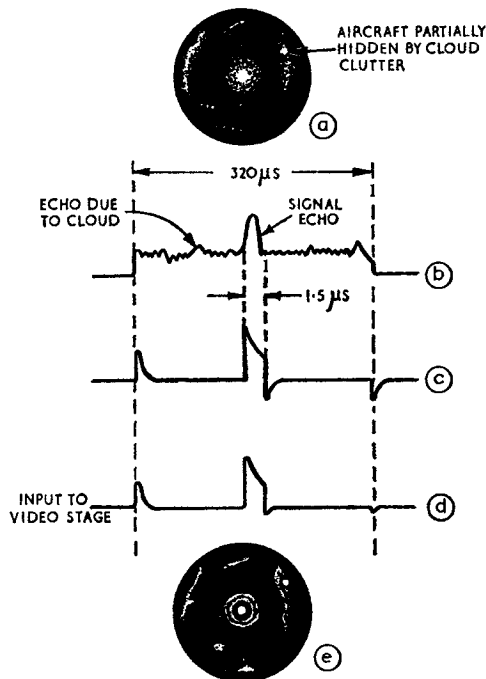


FIG. 3. EFFECT OF FTC CIRCUIT

Gated AGC

The fast time constant circuit is of value in a receiver only so long as the preceding i.f. amplifiers do not saturate. Overloading can be prevented by turning down the gain. This can be achieved manually, as we have seen, but normally some quick-acting form of automatic gain control is also required. Automatic and rapid variation in gain is necessary to ensure that strong echoes do not cause weak echoes at slightly greater range to be suppressed.

In *tracking* radars, the a.g.c. circuit can be caused to operate only at those times when the desired signal is passing through the receiver. In other words, the a.g.c. circuit can be 'gated' to coincide in time with the selected signal. During this time the a.g.c. circuit becomes operative to increase the gain of the receiver. At all other times the gain of the receiver is kept at a low value so that clutter effects are practically eliminated.

The basic outline of the system is illustrated in Fig. 4. A strobe marker is moved along the display by the operator until it coincides with the selected echo. The circuit which produces the strobe marker pulse also produces a gating pulse coincident in time with the strobe pulse.

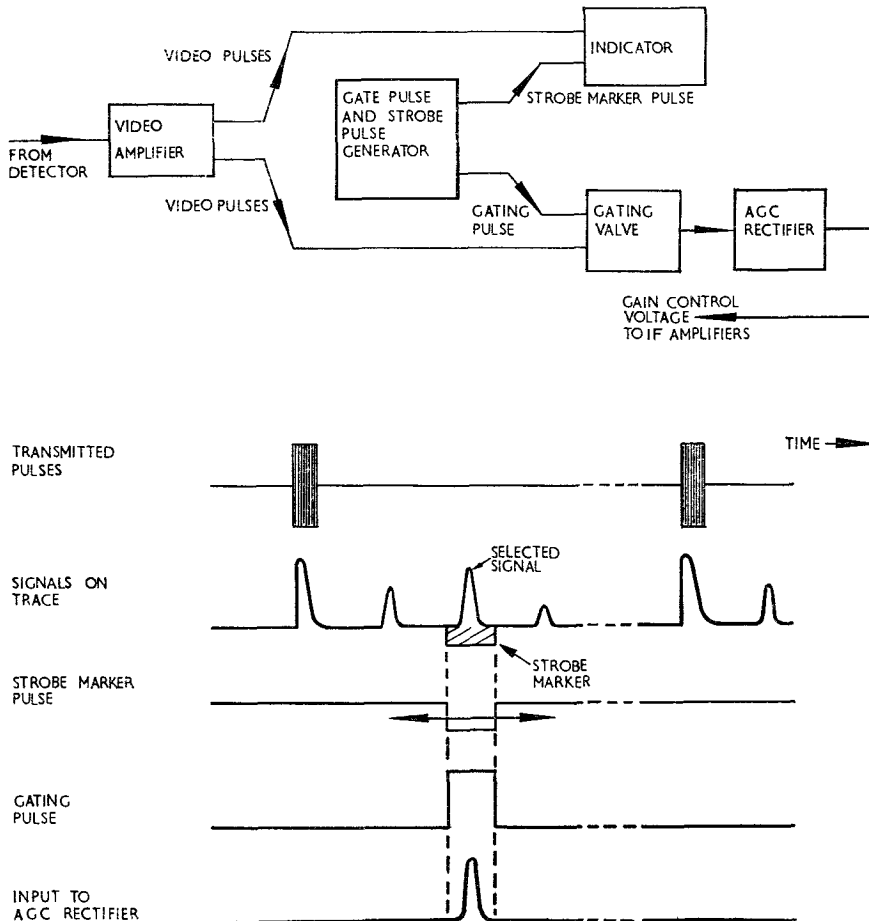


FIG 4. GATED AGC, OUTLINE OF OPERATION

This opens the gate to allow the video output to be applied to the a.g.c. rectifier. The a.g.c. system then applies a control voltage to the i.f. amplifiers to *increase* their gain during the time that the selected signal is being received.

This system applies only to *tracking* radars in which only one target at a time is being examined.

Instantaneous AGC (IAGC)

This is a form of a.g.c. applicable to *search* radars. It is similar to the usual a.g.c. system discussed in Part 1B of these notes (p. 416), except that it is *much faster*. Signals appearing at

the output of the final i.f. stage are rectified by the negative output detector, the output of which is fed to a short time constant CR filter. The output of the filter consists of a negative d.c. voltage which is proportional to the amplitude of the i.f. output and this is applied as a control bias to the grid of the preceding i.f. amplifier (Fig. 5).

The time constant of the CR filter is usually a few pulse durations. The filter therefore acts as a *short CR* to long pulses, such as those from extended clutter targets, and as a *medium CR* to the required signal pulses. Thus the gain of the receiver is immediately reduced when a long clutter pulse is being received but it does not change on receipt of the short duration signal pulse. The echo pulse through the receiver therefore suffers little attenuation but extended clutter pulses are considerably reduced.

Since modulated or unmodulated c.w. acts in much the same way as a 'long pulse', i.a.g.c. is also effective in reducing interference from such sources.

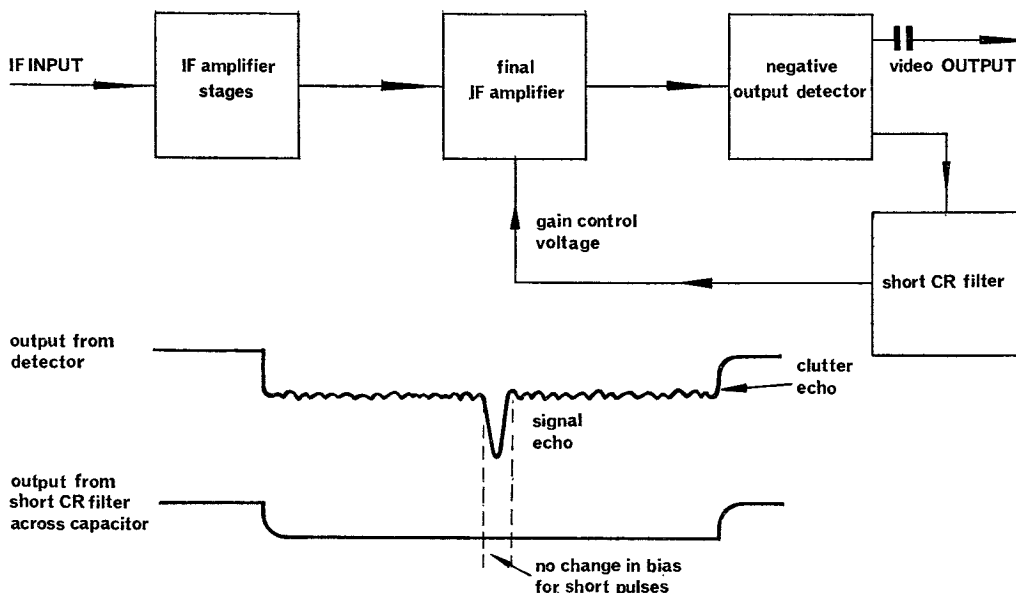


FIG 5. BLOCK DIAGRAM OF IAGC SYSTEM

Usually only one or two stages are controlled by i.a.g.c. because of the danger of instability in the short time constant feedback loop. Since the final i.f. stage will be the first to saturate (signal greatest at this stage) it is to this stage that i.a.g.c. is normally applied.

Anti-clutter Gain Control or Swept Gain

In a ground-based search radar the strongest clutter echoes are from objects immediately surrounding the radar aerial. As the range from the radar increases, the clutter echoes become weaker. On a basic p.p.i. display these effects produce a confused area at the centre of the screen and if the ground clutter is strong enough it can saturate the receiver. The recovery time of the receiver is such that close-range targets are then missed.

Various methods have been adopted in an attempt to reduce strong ground (or sea) clutter. The radar aerial may be tilted slightly upwards to reduce reflections from the ground. Unfortunately, this method also reduces low-angle coverage and is therefore not always suitable. In some cases a radar *diffraction* screen is placed round a radar site to reduce unwanted ground echoes. A diffraction screen, properly sited and constructed, reduces ground clutter with little reduction in low-angle coverage, but it is expensive. We shall see in Section 7 that m.t.i. radar

greatly reduces fixed ground echoes (permanent echoes) in relation to moving targets. But m.t.i. radar does not completely eliminate ground clutter and it also has other limitations.

The use of an *anti-clutter gain control* or *swept gain* circuit is another method that has been used to reduce short-range ground clutter and prevent receiver saturation. It is a relatively simple system. Basically the requirement is that the *gain* of the receiver be *progressively varied* from a low value at the end of the transmitter pulse (i.e. at short ranges) to a high value some time later (i.e. at longer ranges). This definition has led to the use of another name for swept gain, namely *sensitivity time control* (s.t.c.). For targets at close range, when the echo return is strongest, the gain of the receiver is kept at a low value; and for long-range targets, when the echo is weaker, the gain of the receiver is adjusted to a high value. The *rate* at which the gain is made to vary with time can be adjusted by the operator to suit the conditions.

A basic swept gain circuit is illustrated in Fig. 6. The input is a positive-going pulse from the master timing unit. This pulse is coincident in time with that radiated by the transmitter and it has the same pulse duration. It is differentiated by the input network and applied to the diode V_1 . The *amplitude* of the differentiated input voltage is set by R_1 . When V_1 conducts, its cathode voltage V_{1k} rises and lifts V_{2g} . V_{2g} is limited to zero volts by the flow of grid current through R_4 . After the initial rise at A, V_{1k} falls exponentially at a rate determined by C_3R_3 . V_{2g} remains at zero volts until V_{1k} falls to the same level, when V_{2g} also falls exponentially following the fall of V_{1k} . The output from V_2 anode is applied to the grids of the controlled valves in the i.f. amplifier. Thus at instant B, when the transmitter pulse ceases, the gain control voltage is maximum negative (set by R_5 and R_6) and the receiver gain is a *minimum*. The effect of short-range clutter is therefore reduced. Between B and C the gain is being progressively increased at a rate determined by R_3 until, at the range where clutter echoes are considered to be negligible (point C in Fig. 6), the receiver gain is a *maximum* as decided by the setting of the manual gain control R_6 .

A modification to this system is illustrated in Fig. 7. It is known as *r.f. swept gain* because it prevents saturation of the *mixer* and subsequent stages in the receiver. The s.t.c. waveform is produced as described, but instead of applying it to the grids of the controlled i.f. amplifiers it is applied through a triode to the winding of a ferrite phase-shifter inserted in the waveguide crystal mixer bridge. The current through the ferrite winding depends upon whether the triode is conducting or cut off and this, in turn, is controlled by the s.t.c. waveform. From A to B the triode is cut off and the current through the winding is adjusted by R_1 so that the phase-shift introduced by the ferrite rod is such that the effective waveguide length XY is an *even* number of

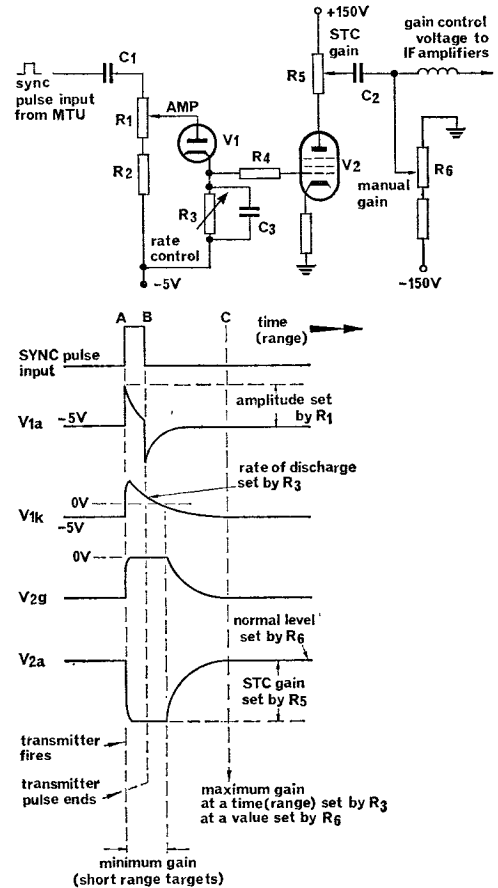


FIG 6. SWEPT GAIN, CIRCUIT AND WAVEFORMS

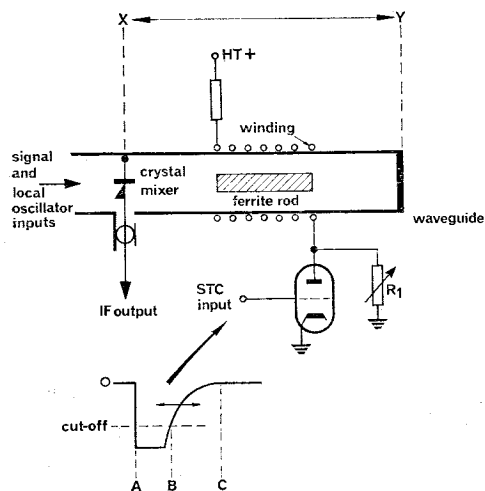


FIG 7. RF SWEPT GAIN

The Logarithmic Receiver

There are certain practical limitations to a swept gain system. The correct adjustment of the rate of change of gain with time depends upon the clutter encountered and some skill is required in selecting the correct setting. If, in addition, the amount of clutter is not the same at all radar bearings, the swept gain setting must be a compromise. Hence ideal swept gain cannot be achieved. However, the functions of the ideal swept gain system may be performed automatically by a receiver with a *logarithmic* input-output characteristic, followed by a differentiating circuit. Such an arrangement is as effective in reducing short-range clutter as a normal *linear* receiver which includes a *perfectly adjusted* swept gain circuit.

We know that a radar receiver giving a linear response to small input signals will usually saturate when the input voltage exceeds a few millivolts. Response to close-range clutter can therefore completely saturate the linear receiver. The result is that targets are lost, *even though their echo may be larger than that due to clutter*. What is required is a receiver giving full amplification to small signals without saturating on strong signals. The logarithmic receiver, which has a response such that the output is the *logarithm* of the input, provides this.

The differences between a linear radar receiver and a logarithmic receiver are in the i.f. amplifier and video stages. The normal i.f. amplifier has a *single* detector at the *end* of the i.f. strip, whereas the logarithmic i.f. amplifier has a detector following *each* i.f. stage in the chain, the outputs of all the detectors being added. Each stage, in addition to feeding the succeeding stage, feeds its own detector and thus makes an *independent contribution* to the final output of the receiver. Fig. 8 shows the outline of the system.

It is seen from Fig. 8 that the outputs of the detectors are added together in a *delay line*. This delay line is correctly terminated at both ends by R_1 and R_2 and it is adjusted so that the delay per section is equal to the delay experienced by the signal pulse in each intervening i.f. amplifier stage. Without this delay line the detector outputs would add 'out of time' in such a way that the final video pulse would have a leading edge consisting of a series of steps; each step would be delayed in time by the time taken for the pulse to pass through an i.f. amplifier stage. The delay line ensures correct addition 'in time' of the detector outputs, and a single-edged output pulse results.

If the input to such a receiver is steadily increased, the *last* stage of the i.f. amplifier is the first to saturate, because the signal amplitude is greatest at this stage. After a further increase in input the last-but-one stage saturates. We can go on in this way, increasing the input and *getting an increased output*, until all the stages are saturated. Thus the range of input voltages

$\frac{\lambda_g}{4}$. Under these conditions the mixer crystal is located in a *null* of the waveguide electric field and so delivers negligible i.f. output. From B to C the triode becomes progressively more conductive and the increased current through the winding adjusts the phase shift introduced by the ferrite rod to give a progressively greater i.f. output from the crystal mixer. At C, the effective waveguide length XY is on *odd* number of $\frac{\lambda_g}{4}$ and the mixer output is a *maximum*. Thus for close-range targets the mixer output is negligible and the output gradually increases to a maximum at the point where clutter is no longer considered to be significant. Saturation by strong ground clutter is prevented.

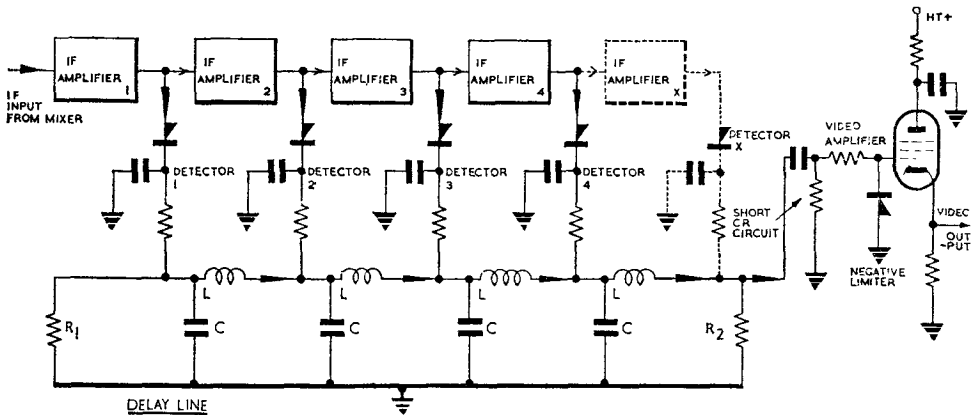


FIG. 8. SCHEMATIC DIAGRAM OF LOGARITHMIC IF AMPLIFIER

over which the receiver output continues to increase is much greater in this type of receiver than in a linear receiver. The *dynamic range* of the logarithmic receiver is large. This can best be shown by an example.

Let us imagine a log i.f. strip consisting of four stages, each with a gain of 100. We shall also assume that when a stage saturates, the output voltage from its detector is 1 volt. Suppose now a signal is fed in to this i.f. strip and that it is just sufficient to saturate stage 4. Then the output from stage 4 is 1 volt. Since each stage has a gain of 100, the *input* to stage 4 (i.e. the output from stage 3) is $(\frac{1}{100})$ volt. Similarly, the output from stage 2 is $(\frac{1}{100})^2$ volt, that from stage 1 is $(\frac{1}{100})^3$ volt, and the input to stage 1 is $(\frac{1}{100})^4$ volt. This is the input voltage which just saturates stage 4. Although the outputs from all the stages add, it may be seen that the small outputs from the stages before the saturated stage are *negligible* compared with the output from the saturated stage.

If the input is now increased 100 times from $(\frac{1}{100})^4$ volt to $(\frac{1}{100})^3$ volt, stage 3 saturates also and this stage now contributes 1 volt to the output. We now have 1 volt due to stage 4, 1 volt due to stage 3, $(\frac{1}{100})$ volt from stage 2, and $(\frac{1}{100})^2$ volt from stage 1. The output is thus approximately 2 volts.

By increasing the input in steps of 100, thus saturating each stage in turn, the result will be as shown in Table 1.

Input (volts)	Output (volts)
$(\frac{1}{100})^4$	1
$(\frac{1}{100})^3$	2
$(\frac{1}{100})^2$	3
$(\frac{1}{100})$	4

TABLE 1. LOGARITHMIC CHARACTERISTICS

By increasing the input voltage over the range $(\frac{1}{100})^4$ to $(\frac{1}{100})$ volt, the output changes from only 1 volt to 4 volts. A logarithmic relationship exists between input and output over the given input range. By increasing the input signal by a 'power' of 4 we have increased the output by only 4 times. The dynamic range of the log receiver is thus very great. It produces an output for very weak input signals but does not easily saturate even for very strong inputs such as might be experienced from close-range clutter objects.

A true logarithmic characteristic cannot be maintained down to zero input. A practical receiver has a *linear-logarithmic* characteristic, the receiver being linear for *small* signal inputs and logarithmic for *larger* signal inputs. In the example given above the log characteristic applies for all signals *larger* than $(\frac{1}{100})^4$ volt. For smaller signals the receiver characteristic is linear.

For successful reduction of clutter effects the logarithmic characteristic should be maintained down to very small signal levels. The figure often quoted is *at least 20 db* below the clutter level.

So far we have described how saturation of the receiver may be prevented by the use of a log i.f. amplifier. To suppress the *effects of the clutter* on the c.r.t. display the output of the delay line in the log amplifier is applied to the video amplifier through a short CR differentiating circuit (an f.t.c. circuit). We have seen that this has little effect on the short-duration signal pulses but it does suppress the longer-duration extended clutter pulses and other forms of 'long pulse' interference.

The output of a log amplifier, when presented on a display, gives a picture which is poor in comparison with that produced by a normal linear receiver. The large amplitude echoes are suppressed by the log characteristic and the differentiation between large and small echoes is less. To make the output more like that of a linear receiver, a video amplifier with the *inverse* of the log characteristic is often used. The video amplifier must be designed to restore the signal-to-noise ratio which has been reduced by the suppression of the large amplitude echoes in the log i.f. amplifier. This may be done by increasing the gain of the video amplifiers and also by biasing the stages further back so that noise is clipped and amplified non-linearly.

Clutter due to Weather Effects

We stated earlier in this book that reflections from fog, rain or snow particles are insignificant at the lower radar frequencies (p. 228). However, since the amount of reflection from an object depends upon the size of the object in relation to the radar *wavelength*, at higher frequencies weather echoes may be strong enough to mask the desired target signals just as any unwanted clutter signal.

It is clear, therefore, that to prevent clutter due to weather effects the radar should be operated at a low frequency. However, this is not always possible. For example, a precision approach radar must provide accurate information and good angle discrimination both in azimuth and in elevation. This can be obtained easily only by the use of high frequency radars; 10,000 Mc/s (3 cm) is a typical frequency band. Thus, where we are using high frequencies, some other means must be provided for reducing weather clutter.

Weather clutter is similar in many respects to ground or sea clutter and several of the techniques used to reduce other clutter echoes have some effect in reducing weather clutter. These techniques include the use of m.t.i. radar (see Section 7), log receivers, swept gain, and f.t.c. circuits. In addition, if the radar operates with a narrow beamwidth and a narrow pulse duration the *amount* of weather clutter energy received back at the radar is reduced.

One of the most common ways of reducing weather clutter is to use *circularly polarized* radiation. We have seen earlier that a circularly polarized wave can be produced from a linearly polarized wave by using a turnstile junction (p. 301) or a quarter-wave plate (p. 325). A circularly polarized wave incident on a spherical object, such as a raindrop, is reflected back to the radar as a circularly polarized wave with the *opposite* sense of rotation. Since the same aerial is used

for both transmitting and receiving, the aerial is not responsive to the opposite sense of rotation and the echo energy due to raindrops and the like will not be accepted by the receiver.

On the other hand a complex target, such as an aircraft, is not a *symmetrical* reflector and it will return *some* energy with the correct polarization for reception. Thus, circular polarization is of considerable value in reducing weather clutter. However, it also reduces the required target echo energy so that arrangements are usually included in the radar to switch to linear polarization

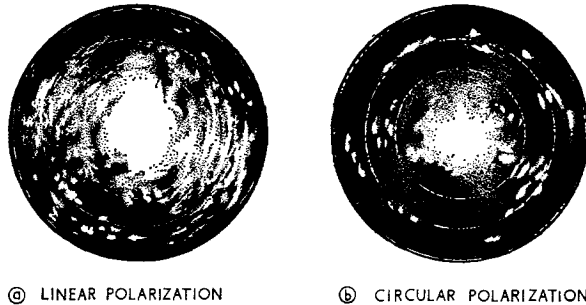


FIG. 9. EFFECT OF CIRCULAR POLARIZATION

whenever weather conditions permit. Where circular polarization is obtained by using a quarter-wave plate we can revert to normal polarization by lowering the plate. Fig. 9 illustrates the effect of circular polarization in conditions of very heavy rain clutter. In Fig. 9a linear polarization is being used; in Fig. 9b, circular polarization. The reduction in weather clutter is clearly seen.

Angels

This is the name given to a form of clutter caused mainly by reflections from birds and insects. For a ground-based radar sited on the coast, clutter due to angels can be almost as great as that due to sea clutter because of the large numbers of seabirds. Although a bird has only a small cross-sectional area compared with that of an aircraft it can return a strong echo at short ranges. A bird at a range of 10 miles can return an echo as strong as that from an aircraft at 100 miles range. And, of course, when birds travel in flocks the clutter caused on the screen by angels can be severe. Fortunately, strong angel echoes occur only at short ranges so that swept gain or the use of a logarithmic receiver can reduce their effect.

INTERFERENCE AND JAMMING

Introduction

In the preceding chapter we considered the effects of clutter on receiver performance and discussed ways in which these effects could be reduced. It will be remembered that clutter echoes are caused by the *radar's own transmissions*. Another class of unwanted signal that can enter the receiver is *interference* from nearby transmitters. This interference may be *unintentional* and originate from 'friendly' communication or radar transmitters; or it may be *intentional* and originate from a 'hostile' *electronic countermeasure* (e.c.m.) transmitter. In this chapter we shall look briefly at both types of interference and mention some of the steps that can be taken to reduce their effects.

Unintentional Interference

In areas where there are many radar and communication transmitters, mutual interference between neighbouring stations can be severe. The interference may be c.w., modulated c.w., or pulse. We shall consider mainly pulse interference such as might be caused by other radars.

The interference from a nearby radar usually consists of a train of r.f. pulses whose frequency may or may not be the same as that of the victim radar. The p.r.f. of the interfering signal will not usually be the same as that of the desired signal so that the interfering pulses will not appear at the same range on each sweep of a radar display. The effect of this on the display is shown in Fig. 1. Because of its appearance this type of interference is called '*railing*'. A similar form of interference is often seen on television.

Mutual interference between radars may be reduced by operating neighbouring radars on different frequencies. However, we know that the band of frequencies occupied by a transmitted pulse is very wide. Thus, interference may be caused in a receiver whose operating frequency is quite different from that of the interfering transmitter.

Radar receivers which operate in conditions of heavy interference are usually well shielded to reduce strong pick-up. Filters may also be included in the receiver input circuits to reduce interfering signals from radars operating on adjacent frequencies. Special circuits which discriminate between the desired and the interfering signals may also be included. Typical of such circuits are the *pulse duration discriminator* (p.d.d.) and the *p.r.f. discriminator* (p.r.f.d.).

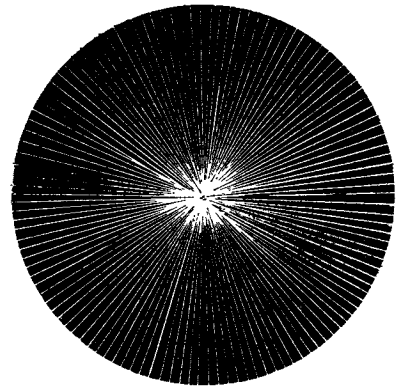


FIG. 1. 'RAILING' CAUSED BY INTERFERENCE

Pulse Duration Discriminator

If the pulse duration of the interference differs from that of the desired signal the *difference* in pulse durations may be used as a means of discrimination between the two signals. A simple block diagram of the system is illustrated in Fig. 2. The input pulse is differentiated and applied to two circuits. In one circuit the pulse is *delayed* by a time equal to the pulse duration of the desired signal. In the other circuit the differentiated pulse is *inverted*. If the input pulse is the desired signal pulse, the trailing edge of waveform 3 will coincide with the leading edge of

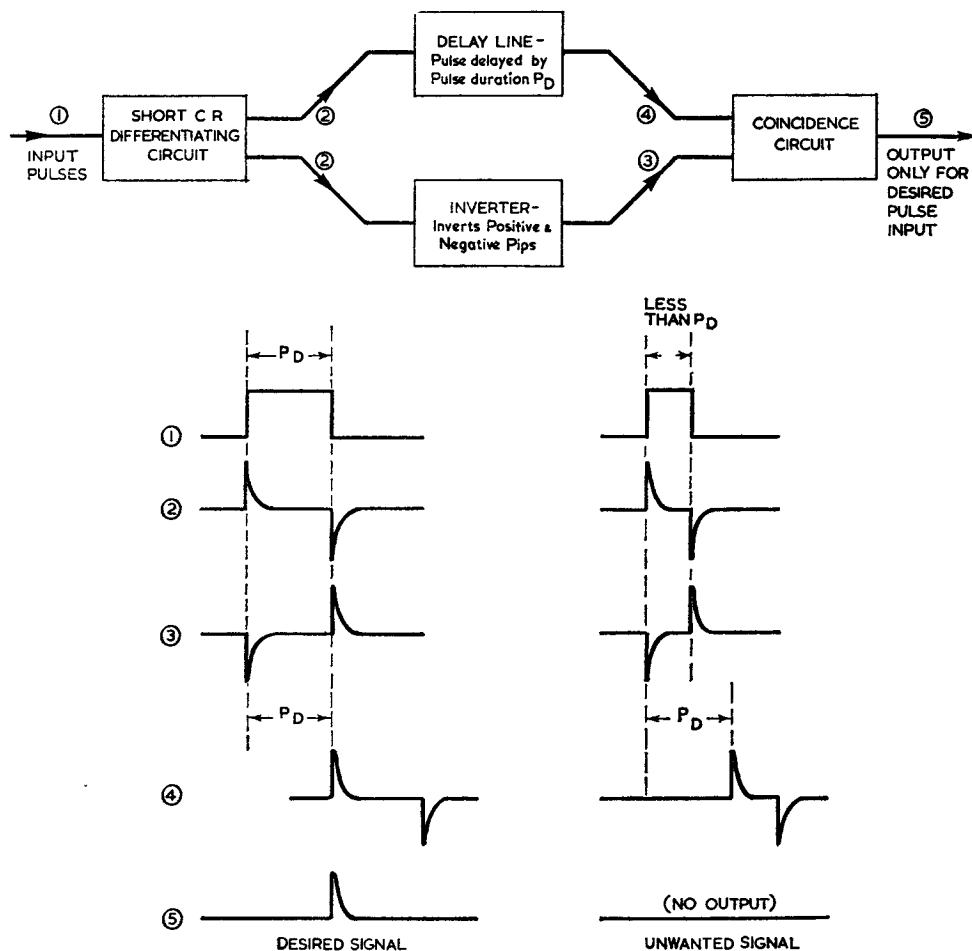


FIG. 2. PRINCIPLE OF PULSE DURATION DISCRIMINATOR

waveform 4. If the input pulse is *not* of the required duration the two pips of the differentiated waveforms will not coincide. The coincidence circuit provides an output only if the two pips shown in Fig. 2 are coincident in time. Thus, only the desired signal pulse provides an output; the interfering pulse is rejected.

PRF Discriminator

We know from previous work that if we apply a train of pulses to an *integrating* circuit the output voltage is related to the *number* of pulses applied in a given time. It is possible to arrange for the integrator to produce maximum output when the input consists of, say, x pulses per second; for inputs whose p.r.f. is lower or higher than this value, the output falls to a low value. Such an integrator in effect 'selects' the required p.r.f.

This is the basis of the p.r.f. discriminator used in the video stages of some radar receivers. A video delay line is used in the integrator and the delay time is adjusted to equal the reciprocal of the desired p.r.f. The integrator then has maximum response to inputs at the desired p.r.f. For interfering signals which have a lower or higher value of p.r.f. the output is very small.

Fig. 3 shows the effect of using a p.r.f. discriminator. In Fig. 3a the railing is so intense that desired targets are obscured. In Fig. 3b the railing has been effectively suppressed by the use of a p.r.f. discriminator.

Other Methods of Reducing Interference

Interference among radars located *at the same station* can be reduced if they are all operated at the same p.r.f. and if they are synchronized to fire simultaneously. This is the method adopted at many RAF radar stations. The master timing unit generates a master triggering pulse which causes all radars on the site to operate simultaneously.

Interfering signals over a wide band of frequencies can be reduced by subtracting the video output of the radar receiver from the output of an auxiliary receiver *tuned to a different frequency*. The output from the auxiliary receiver consists only of interference; that from the radar receiver consists of the desired signal plus the interference. By subtracting the video outputs from the two receivers, only the desired signal (plus noise) remains.

ECM and ECCM

In war, a radar may be subjected to deliberate interference or *jamming* with the object of reducing the effectiveness of the radar. The methods used to produce jamming are called *electronic countermeasures* (e.c.m.). To get sufficiently close to the radar which is to be jammed, the e.c.m. equipment is usually carried in a jamming aircraft.

To allow the radar to operate under conditions of jamming, the effects of the interference must be counteracted.

The techniques used to provide release from jamming are known as *electronic counter-counter-measures* (e.c.c.m.).

There are two classes of e.c.m.: those intended to cause confusion, and those intended to cause deception.

a. Confusion e.c.m. This masks the display of targets on the radar screen by producing effects similar to those produced by ground or sea clutter, but to a much greater extent. Effective confusion e.c.m. completely obliterates the radar screen. Its effect can be partially off-set by good design of the radar receiver. We have seen in an earlier chapter that a radar receiver designed as a *matched filter* has an optimum bandwidth at which the signal-to-noise ratio is maximum (p. 405). A matched filter radar receiver rejects a large amount of the jamming. Other points of good general design, which improve the noise factor of the receiver, are the use of low-noise circuits and careful screening.

b. Deception e.c.m. This is designed to produce false echoes which appear on the radar screen as though they were real targets. Since deception echoes can be of about the same power as real echoes, the power output requirements of deception jammers are low. To combat deception e.c.m. special techniques are required at the receiver.

Both confusion and deception e.c.m. may be produced by either *passive* devices or *active* devices. Passive devices do not radiate of their own accord; they merely re-radiate the incident radar pulse. Active devices contain a transmitter which radiates an interfering signal. We thus have four possible e.c.m. sources: passive confusion e.c.m., active confusion e.c.m., passive deception e.c.m., and active deception e.c.m. We shall look at each in turn.

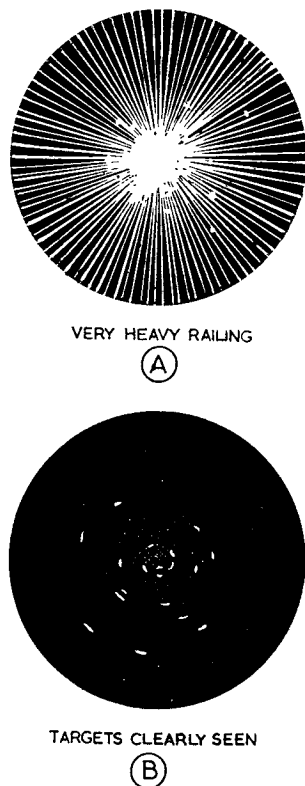


FIG. 3. EFFECT OF PRF DISCRIMINATION

Passive Confusion ECM

One of the earliest forms of e.c.m. used against radar was *window*. Window consists of a mass of tinfoil strips, each strip acting as a dipole reflector. If window is continuously released from a jamming aircraft it forms a 'smoke-screen' behind which following aircraft can fly undetected. On the screen of the ground radar, window produces a confused area of extended clutter within which real targets are difficult to detect. Since confusion e.c.m. produced by window has an effect on the c.r.t. screen similar to that produced by normal clutter it may be reduced by the methods considered earlier.

Active Confusion ECM

Simple c.w. jamming. A simple method of jamming is for an e.c.m. aircraft to transmit a c.w. signal at the frequency of the radar being jammed. Fortunately, as we have seen, c.w. interference is easily suppressed by the insertion of an f.t.c. filter circuit between the detector and video amplifier stages (p. 428). Alternatively, it can be reduced if the receiver uses i.a.g.c. (p. 430).

Narrow-band or 'spot' noise jamming. A better method of jamming is to modulate the e.c.m. transmitter with a *noise* signal. In spot jamming, the jammer radiates a noise-modulated signal at the frequency of the radar. The bandwidth of the radiated noise energy is sufficient only to cover the radar receiver bandwidth so that large amounts of noise power enter the receiver circuits. The intended effect is to cover the whole screen with clutter and make it impossible for the operator to detect targets. If the spot jammer power is large enough, the entire display can be swamped.

The main counter-countermeasure takes advantage of the fact that the spot jammer concentrates a large noise power over a *narrow bandwidth*. Thus, by changing the radar frequency, spot jamming may be avoided. To do this a *tunable* radar is needed; the radar frequency must be changed rapidly so that the jammer does not have time to follow. An ideal arrangement would be to change the frequency from pulse to pulse. An auxiliary receiver could be used to monitor the whole band continuously and so provide information on which frequencies were free from jamming.

Another effective counter-countermeasure is to reduce the *side lobes* of the radar aerial as much as possible. Maximum confusion will be produced if the noise jamming power is sufficient to enter the aerial side lobes. If the side lobes are reduced so that the jamming is picked up only by the main lobe the effect of the jammer is considerably reduced (Fig. 4a). For a less efficient aerial, some jamming power enters by the side lobes (Fig. 4b); and for an aerial with large side lobes almost complete jamming results (Fig. 4c). Reduction of the side lobes is important for another reason: Fig. 4a shows the approximate bearing of the jammer; by triangulation with two or more similar radars the approximate position of the target (jammer) can be calculated.

Wide-band or 'barrage' noise jamming. A wide-band jammer radiates noise over a wide band of frequencies. The bandwidth is usually such that it covers the entire frequency range of a particular class of radar, e.g. a 3 cm radar. Hence the use of a tunable radar is no defence against barrage jamming.

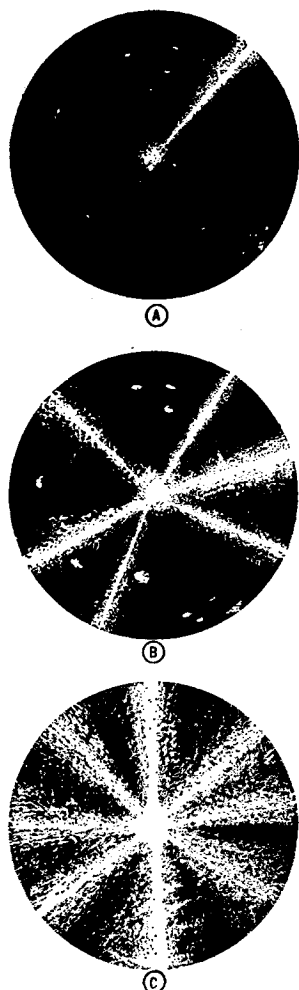


FIG. 4. IMPORTANCE OF REDUCING AERIAL SIDE LOBES

One counter-countermeasure is to have several radars in different frequency bands, e.g. u.h.f., 3 cm, 10 cm. This will force the e.c.m. aircraft to carry more jammers than may be convenient.

Although the barrage jammer is more effective than the spot jammer in many respects, it is less effective in one respect: since the available power of the barrage jammer is spaced over a very wide frequency range the *amount* of noise power entering the receiver passband is less than that from a narrow-band spot jammer having the same power output.

A barrage jammer may be obtained by amplifying the noise output of a wideband receiver and using this to modulate the jammer. Another type of barrage jammer uses a carcinotron (p. 275), *frequency-modulated by noise* to produce a random sweep in frequency. This is known as *sweep-through* jamming. To be effective as a wide-band jammer, the sweep in frequency must be large compared with the radar receiver bandwidth. Each time the carrier from the jammer sweeps through the receiver passband, a pulse is produced. The pulses are *randomly* spaced because the carrier sweep is random. For maximum confusion the barrage jammer is adjusted in such a way that the time taken for the carrier to sweep through the receiver bandwidth is equal to the time taken by the receiver to respond to an echo. A single sweep-through jammer can simultaneously jam many different radars operating at different frequencies within a given band.

Wide-band jamming can be reduced by using a matched filter receiver and by designing an aerial with small side lobes. Where the side lobes are small, the position of the jamming aircraft can be plotted by triangulation using the main lobes from two or more ground radars. The other counter-countermeasure is to have several radars in *different bands*.

Passive Deception ECM

Window. We have seen how window can be used to produce confusion e.c.m. It can also be used to provide deception. If window is released from an aircraft *in a bundle*, the bundle slowly separates and produces an echo at the ground radar similar to that produced by a large aircraft. The release of several bundles in succession produces deception by causing additional 'targets' to be seen on the radar screen.

The counter-countermeasure lies in the fact that window is quickly recognised as not being a true target since its speed is much less than that of an aircraft. The necessary discrimination is performed either by an operator or by the use of m.t.i. radar.

Decoys. A decoy is a small unmanned aircraft which can be launched from the attacking aircraft outside the range of the radar. When picked up by the ground radar, the decoy produces deception echoes. The decoy can be made to give an echo similar to that of a large aircraft by fitting it with reflectors. The remedy is to destroy every hostile target which appears on the screen, including targets which are known to be decoys (the decoy may carry a bomb).

Rotating reflectors. Tracking radars which use a conical scan can be deceived by a rotating reflector mounted on the aircraft. If the reflector

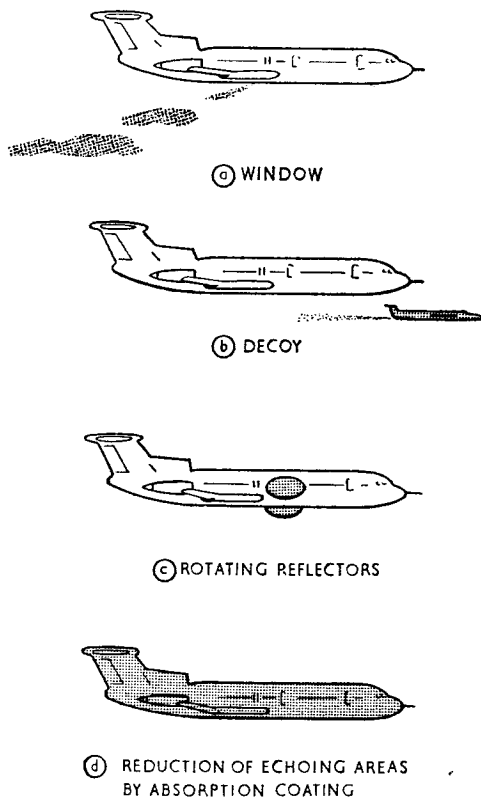


FIG. 5. PASSIVE DECEPTION ECM METHODS

is rotating at the correct speed, it produces amplitude modulation of the echo signal returning to the radar receiver. This may be wrongly interpreted as a change of target aircraft position. The remedy here is to use the *monopulse* tracking system (p. 337).

Reduction of radar echoing area. The echoing area is the effective cross-section of a target as seen by the incident pulse from the radar. It can be reduced, thus deceiving the ground radar, by the use of 'radar-absorbing' materials.

a. Interference absorbers. The jamming aircraft may be given a 'coating' of iron particles, the coat having a thickness of $\frac{\lambda}{4}$ at the radar wavelength. This coating has the effect that energy reflected from the front surface is largely cancelled by that which enters the coating and is reflected from the back surface. Because of the dependence of the thickness of the coat on wavelength, interference absorption has a narrow bandwidth.

b. Attenuation absorbers. The jamming aircraft may be coated with several thick layers of conducting material, the layers being separated from each other by a plastic sheet. The layers are of such material that their conductivity gradually increases with depth. The reflection from the front surface is small and as the wave penetrates the layers it is rapidly attenuated. This method requires a thicker coating than for the interference absorber but it is largely frequency-independent and has, therefore, a wide bandwidth.

Active Deception ECM

Repeater jamming. The jamming aircraft may carry a low-power transmitter which is triggered by the incident pulse from the ground radar. By delaying the received pulse and re-transmitting a slightly different pulse after a short time interval, false echoes are produced. The false echoes appearing on the ground radar screen are at a range and bearing different from that of the real target.

Transponder repeater. This plays back a *stored replica* of the received radar pulse shortly after the transponder has been triggered by the ground radar. By suitable arrangement of the circuit, the transponder can be silent when illuminated by the main radar beam and made to respond only to the side lobes. By doing this it creates false echoes which are vastly different in position from the true target.

Range-gate stealer. This is a repeater jammer which causes a pulsed tracking radar to break its lock on a target. We saw in Section 2 (p. 215) that a tracking radar generates a pair of range gates, between which the target echo is aligned. As the target moves, the range gates automatically move also to give continuous tracking in range. The range-gate stealer starts by transmitting a pulse in synchronism with the real echo, thus strengthening the real echo. If the stealer pulse is now slowly shifted in time, and is stronger than the true echo, the ground tracking radar will lock on to the false pulse and ignore the much weaker echo from the true target. The result is a completely false indication of range of the target.

Velocity gate stealer. CW radar may also be used in the tracking role (see Section 7). To deceive a c.w. radar which is relying on the doppler shift in frequency to track a target, a repeater jammer may operate as a velocity-gate stealer. This transmits a signal which gives false information on the target speed; it may even indicate that the received echo is from a stationary target.

Repeater-type jamming may be reduced by transmitting a pulse with a form of identification difficult for the jammer to imitate. The jammer can also be rendered ineffective if the ground radar switches to different values of pulse duration or p.r.f. or to a different polarization. The use of monopulse and aeriels with small side lobes are also useful counter-countermeasures.

Summary

Table 1 summarizes the various methods of e.c.m. and e.c.c.m. discussed in this chapter.

	CONFUSION		DECEPTION	
	ECM	ECCM	ECM	ECCM
PASSIVE	(a) Window	(a) Good receiver design (matched filter); f.t.c. and i.a.g.c. circuits.	(a) Window (b) Decoys (c) Rotating reflectors (d) Reduction of echoing area.	(a) MTI radar. (b) Destroy all targets. (c) Monopulse system. (d) Good receiver design.
ACTIVE	(a) CW jammer	(a) Good receiver design; f.t.c. and i.a.g.c.	(a) Repeater jammer	(a) Use a pulse difficult to imitate; change pulse duration, p.r.f., or polarization; small aerial side lobes.
	(b) Spot noise jammer	(b) Matched filter receiver; tunable radar; small aerial side lobes.	(b) Transponder jammer.	(b) As for (a).
	(c) Barrage noise jammer.	(c) Matched filter receiver; several radars in different bands; small aerial side lobes.	(c) Range-gate stealer. (d) Velocity-gate stealer.	(c) Monopulse system. (d) Monopulse system.

TABLE 1. SUMMARY OF ECM AND ECCM METHODS

SECTION 7

CW RADAR

Chapter		Page
1	CW Ground Radar	445
2	FMCW Radar Altimeters	465
3	Airborne Doppler Principles	477
4	Moving-target Indication (m.t.i.) Radar	499

CHAPTER 1

C.W. GROUND RADAR

Introduction

A basic outline of c.w. radar is given in Section 1 of this book (see p.53). There we saw that if there is *relative motion* between a target and a radar, then the frequency of the energy received back at the radar differs from that radiated. This *difference* in frequency is known as the *Doppler shift*. If the target is *approaching* the radar, the received frequency is *higher* than that transmitted; for a *receding* target the received frequency is *lower*.

In this section we shall examine c.w. radar in more detail. In this first chapter we shall consider c.w. radar principles and their application to ground radars; in Chapter 2 we shall look at the application of frequency-modulated c.w. radar to altimeters; in Chapter 3 we shall deal with the use of Doppler for aircraft navigation systems; and in Chapter 4 we combine pulse and Doppler principles to see how they are used in moving-target indication (m.t.i.) systems.

The purpose of any radar is to detect a signal returned from a target and to extract the required information from that signal. This information may include the bearing and elevation of the target, and its range or relative velocity (or both). For many years the normal method of detecting targets has been by pulsed radar. Although adequate for many purposes, pulsed radar has limitations. It cannot easily distinguish between fixed and moving objects, and the fixed objects produce clutter which may completely mask the required moving target. Measurement of the relative velocity of a target by pulsed radar is also difficult. A c.w. radar system can ignore stationary objects and so overcomes these, and other, limitations of the pulsed radar.

The Doppler Effect

In c.w. radar the transmitter is operating continuously and one of the problems in detection is to separate the weak echo from the strong transmitted signal. The two can be separated fairly easily if the signal received back from the target differs in frequency from that being transmitted. As we have seen, this occurs *automatically* in c.w. radar by virtue of the Doppler effect.

Fig. 1 shows a transmitter and a receiver separated by a *variable* distance d . The transmitter is radiating a signal of the form $\cos \omega_T t$, where the suffix T refers to the transmitter. The *phase* of the radiated signal changes by 2π radians for each wavelength travelled, and over the distance d

the phase change is $\frac{d}{\lambda_T} \cdot 2\pi$ radians, where λ_T is the wavelength of the transmitted signal. The received signal can therefore be represented by $\cos \left(\omega_T t - 2\pi \frac{d}{\lambda_T} \right)$.

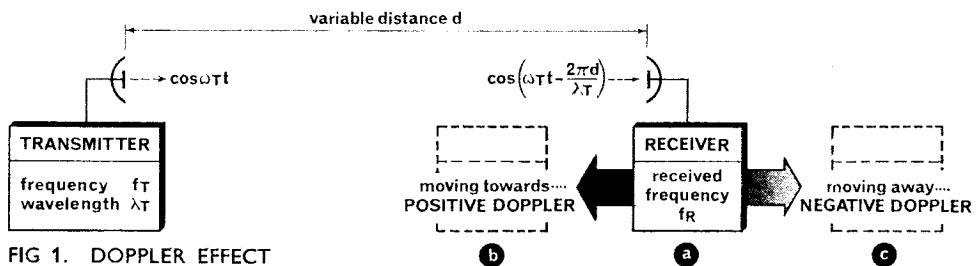


FIG 1. DOPPLER EFFECT

The angular velocity of the received signal (ω_R) is given by the *rate of change* of its phase with respect to time. Thus:

$\omega_R = \omega_T - 2\pi \frac{\dot{d}}{\lambda_T}$, where $\dot{d}$ is the rate of change of distance with time. We shall refer to $\dot{d}$ as

relative velocity and give it the symbol v_r . Thus:

$$\omega_R = \omega_T - 2\pi \frac{v_r}{\lambda_T}$$

$$f_R = f_T - \frac{v_r}{\lambda_T}, \text{ since } f = \frac{\omega}{2\pi}$$

In this expression f_R is the received frequency and f_T the transmitted frequency. The *difference* in frequency between f_R and f_T is $\frac{v_r}{\lambda_T}$ and is known as the *Doppler shift* f_D . In general, for the case illustrated in Fig. 1:

$$f_R = f_T \pm \frac{v_r}{\lambda_T}.$$

The positive sign is taken for distance d decreasing and the negative sign for d increasing. If the receiver is *approaching* the transmitter (Fig. 1b) the received frequency f_r is *higher* than the transmitted frequency f_T by $\frac{v_r}{\lambda_T}$ and the Doppler shift f_D has a *positive* sign. If the receiver is *moving away* from the transmitter (Fig 1c), f_R is *lower* than f_T by $\frac{v_r}{\lambda_T}$, and f_D has a *negative* sign.

So far we have considered the effect of a moving *receiver*. If the receiver were stationary and we caused the *transmitter* to move we should get similar results to those described above: for a transmitter moving at velocity v_r *towards* a stationary receiver the received signal is *higher* than the transmitted signal by an amount equal to the Doppler shift; if the transmitter is *moving away* from the stationary receiver, f_R is *less* than f_T by the Doppler shift f_D .

In primary radar, these two effects are normally combined because the target acts both as a receiver and as a transmitter. The energy *received* at a moving target, is Doppler-shifted in frequency by $\frac{v_r}{\lambda_T}$; but some of this energy is *re-radiated* by the target, now acting as a moving transmitter, and a further Doppler shift of $\frac{v_r}{\lambda_T}$ occurs. Thus, the *total shift* in frequency between the energy at frequency f_T radiated from the radar and that of frequency f_R received back at the radar is $(\frac{v_r}{\lambda_T} + \frac{v_r}{\lambda_T})$ or $2 \frac{v_r}{\lambda_T}$. Hence in primary radar:—

$$\text{Doppler shift } f_D = \pm 2 \frac{v_r}{\lambda_T}$$

$$f_D = \pm 2 \frac{v_r f_T}{c} \quad (\text{since } \frac{1}{\lambda_T} = \frac{f_T}{c}).$$

The positive and negative signs indicate an approaching or receding target. The units used in this expression must be consistent, i.e. if the relative velocity is given in m.p.h., the velocity of light c must also be given in m.p.h.

It may be useful to note that the Doppler shift in frequency is approximately 30 Hz per 100 MHz radiated frequency per 100 m.p.h. relative velocity. Thus, if we are using a 10,000 MHz X-band c.w. radar, a target approaching with a relative velocity of 300 m.p.h. will produce a 'positive' Doppler shift of approximately 9,000 Hz ($f_R > f_T$). For a target moving away from this radar at a relative velocity of 1,000 m.p.h., a 'negative' Doppler shift of 30,000 Hz is produced

($f_R < f_T$). If we measure the Doppler shift f_D , this gives a direct measurement of the relative velocity v_r of a target, since the other factors, c and f_T are constant. Hence, using the figures above, a beat note of 9,000 c/s from a 10,000 Mc/s radar indicates a target relative velocity of 300 m.p.h.

Effect of Target Heading

So far we have assumed that the target was moving along the *sight-line*, i.e. moving in a line directly towards or away from the radar. If the target is not moving along the sight-line its velocity *relative to the radar* is less than the actual velocity of the target and the resultant Doppler shift is also less. It may be seen that the relative (or radial) velocity of a target is the component of actual velocity taken along the sight-line.

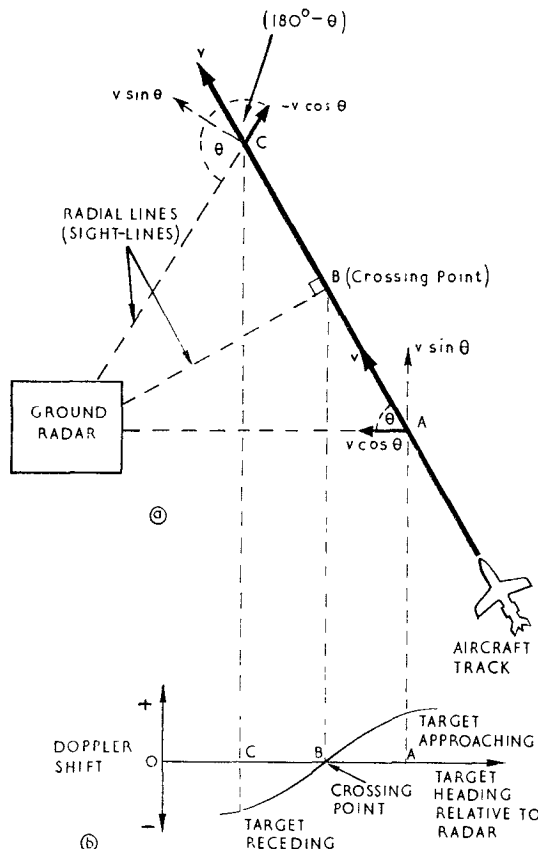


FIG. 2. EFFECT OF TARGET HEADING ON RELATIVE VELOCITY

Fig. 2a shows an aircraft flying on a heading ABC past a radar installation. At A the aircraft track makes angle θ to the radial line (sight-line) joining the radar and the target. The vector representing aircraft speed and track can be split into two component vectors, at 90° to each other. The component of velocity *towards* the radar is the relative velocity $v_r = v \cos \theta$, where v is the aircraft speed. The Doppler shift produced by this relative velocity has a 'positive' sign and has a magnitude of:

$$2 f_T \frac{v \cos \theta}{c}$$

When the aircraft reaches B it is, at this instant, flying *at right angles* to the sight line and has zero velocity *relative to the radar*. The Doppler shift is now zero.

When the aircraft is at C it has a relative velocity *away* from the radar of $v \cos \theta$. The Doppler shift is now 'negative' and has the same magnitude as above.

Hence, when a target flies across the path of a radar, as in Fig. 2a, the Doppler shift varies from positive through zero to negative values. The zero point is called the *crossing point* (Fig. 2b).

Simple CW Radar System

In Fig. 3 the transmitter radiates a c.w. signal of frequency f_T . For a target

moving with relative velocity v_r , the received signal is shifted from f_T by $\pm f_D$, where f_D is the Doppler shift. The received signal at $f_T \pm f_D$ beats in the mixer with a portion of the transmitter output at f_T to produce the Doppler shift of f_D . With this simple system the 'sign' of the Doppler shift is lost in the mixing process. The signal at the beat frequency f_D is amplified and applied to the indicator. The indicator may be a normal frequency meter which measures f_D but is calibrated in terms of the radial velocity v_r (since v_r is proportional to f_D).

Although the transmitter and receiver in Fig. 3 are shown with separate aerials, it is possible in low-power systems to use a *common aerial*. As we noted earlier, the necessary isolation between transmitter and receiver is the separation in frequency $\pm f_D$ caused by the Doppler effect. However, additional isolation is usually necessary to prevent damage to, and saturation of, the receiver from the relatively high-power output of the transmitter. A TR switch, such as is used in a pulsed radar, cannot be used here because in a c.w. system both the transmitter and the receiver are operating continuously. For *low-power* radars (up to ten watts or so), a common aerial may be used and additional isolation between transmitter and receiver can be obtained by inserting a hybrid junction in the feed to and from the common aerial. The hybrid junction may be a magic-T, rat-race, or short-slot coupler (see p 408). Other possibilities for low-power radars include the use of a ferrite *isolator* or circulator, and separate polarisations for transmission and reception. For *high-power* c.w. radars (above several hundreds of watts) it is usually necessary to use separate aerials for transmitter and receiver to obtain adequate isolation. This is the usual practice for ground c.w. radars. The receiver aerial is placed slightly behind the transmitter aerial so that 'spill-over' from the transmitter to the receiver is small.

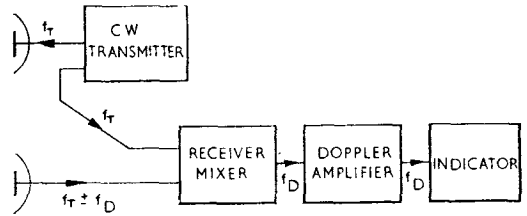


FIG. 3. SIMPLE CW RADAR SYSTEM

Superheterodyne CW Receivers

Fig. 4 shows the block diagram of a superheterodyne receiver suitable for the reception of Doppler signals. The local oscillator is unusual in that it operates at the same frequency as the i.f. of the receiver. The output from this oscillator mixes with a portion of the transmitted output

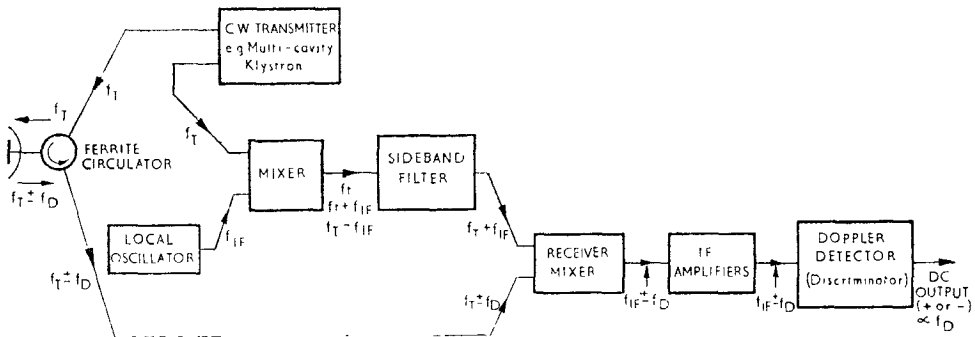


FIG. 4. SIDEBAND SUPERHETERODYNE RECEIVER

at f_T , and the mixer output contains components at f_T , $f_T + f_{IF}$, and $f_T - f_{IF}$. One of the 'sideband' components (in this case $f_T + f_{IF}$) is selected and passed by a filter to the mixer in the receiver. There it combines with the received signal at $f_T \pm f_D$ to produce an output at $f_{IF} \pm f_D$. This signal is amplified in the i.f. amplifiers and then detected to produce a d.c. output of polarity and magnitude proportional to the Doppler shift.

The bandwidth of the i.f. amplifiers handling the signal $f_{IF} \pm f_D$ must be wide enough to pass the expected range of Doppler frequencies. A wide-band amplifier introduces additional noise into the receiver, because of the dependence of noise on bandwidth, and the receiver sensitivity decreases. To counteract this, the receiver may be designed as a *matched filter* (see p 405). However, to do this, the Doppler frequency shift must be known accurately and this is not always possible.

When the Doppler shift is known to lie somewhere within a certain *band* of frequencies, a *bank* of narrow-band i.f. filters, spaced to cover this band, can be used to measure the actual Doppler shift. Crystal or mechanical filters may be used to provide the narrow bandwidth. Fig. 5 illustrates a simple example. The i.f. is 1 Mc/s and the expected Doppler shift lies within the range ± 50 kc/s. Thus the total required bandwidth is 100 kc/s, covering the band 950 — 1,050 kc/s. One hundred narrow-band filters spaced throughout this band may be used, each having a bandwidth of 1 kc/s measured at the half-power (3db) points. The responses of five of the filters in the immediate vicinity of the centre i.f. are illustrated in Fig. 5b. Filter 1 is centred on 998 kc/s and will pass these Doppler shifts which lie 2.5 kc/s to 1.5 kc/s below the centre i.f. Filter 2 is centred on 999 kc/s and will pass those Doppler shifts which lie 1.5 kc/s to 0.5 kc/s below the centre i.f.; and so on for the other filters.

The entire bandwidth is covered in this way and by noting which filter provides the output the Doppler shift can be measured and used to indicate the relative velocity of the target. By reducing the bandwidth of each filter and inserting more filters to cover the required bandwidth the accuracy of measurement is increased. The *sign* of the Doppler shift is also preserved; if the filter which provides the output is *below* the centre i.f. in frequency the target is *moving away* from the radar; if it is *above* the i.f., the target is *approaching*. The output of each filter is applied to a detector which provides a d.c. output for operating the indicator.

It should be noted here that although the i.f. bandwidth of a Doppler receiver must be wide enough to pass the expected range of Doppler frequencies, the bandwidth is measured in kc/s compared with Mc/s for a pulsed radar (see p 65).

In the system described, the sign and frequency of the Doppler shift are measured in the *i.f. stages* of the receiver. For accurate measurement of the Doppler shift each filter would have to be designed with a very narrow passband (often of the order of 100 c/s). This is very difficult to achieve when the filters are operating at or near the intermediate frequency. It is often easier

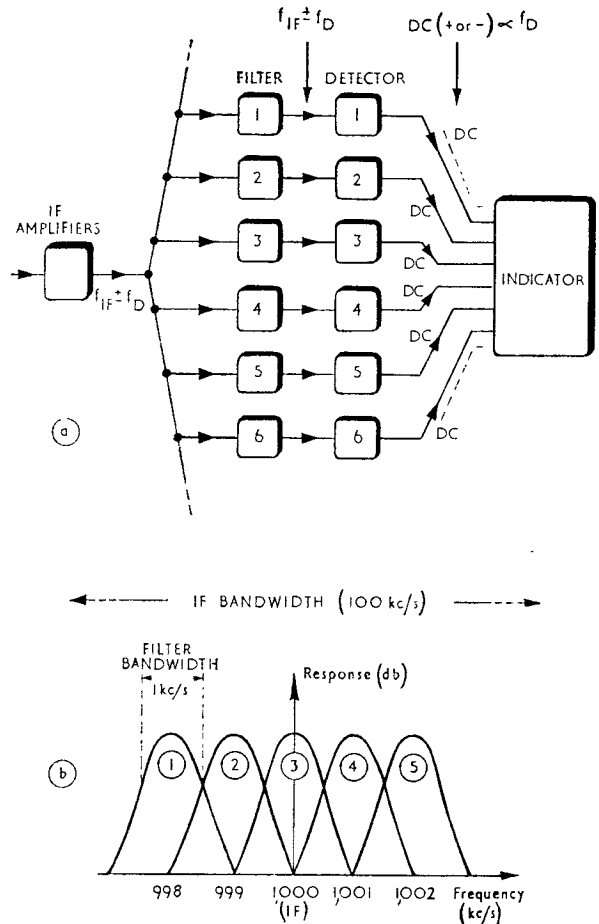


FIG. 5. USE OF BANK OF IF FILTERS FOR DOPPLER MEASUREMENT

and more practicable to extract the relatively low-frequency Doppler shift *before* measuring its frequency and sign. Practical methods of achieving this are described later.

Another type of superheterodyne c.w. receiver which has found considerable application is illustrated in Fig. 6. It is often referred to as a 'signal-following' receiver because an a.f.c. circuit causes the local oscillator frequency to vary in sympathy with any variations in the transmitted signal frequency.

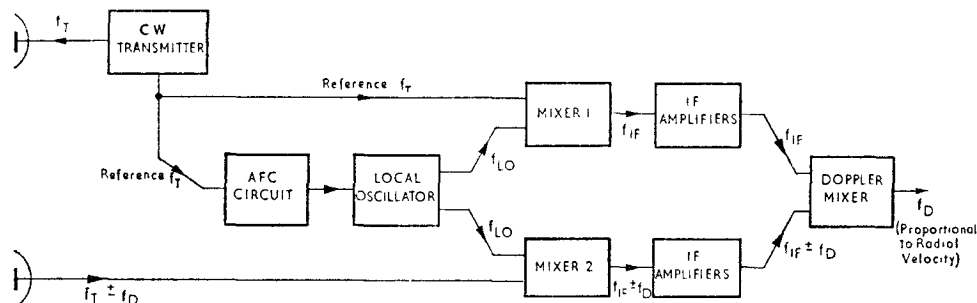


FIG. 6. SIGNAL-FOLLOWING SUPERHETERODYNE RECEIVER

In this receiver both the reference signal at the transmitter frequency f_T and the received signal at $f_T \pm f_D$ are reduced to a suitable intermediate frequency by the use of a common local oscillator. Mixer 1 combines the local oscillator output and the reference signal at f_T to produce an output at f_{IF} . Mixer 2 combines the local oscillator output and the received signal at $f_T \pm f_D$ to produce an output at $f_{IF} \pm f_D$. These two i.f. signals are amplified separately and applied together to the Doppler mixer, where the difference term f_D is extracted. This is the required Doppler shift which is proportional to the relative velocity of the target.

If we look inside the box labelled 'Doppler mixer' in Fig. 6 we shall find that it contains *phase-sensitive detectors*. These are used to compare the *phase* of the Doppler-shifted received signal at $f_{IF} \pm f_D$ with that of the reference signal f_{IF} derived from the transmitter. The output of the phase-sensitive detectors is the required Doppler shift. There are many forms of phase-sensitive detectors. We shall now consider one of the simplest.

Phase-sensitive Detector

Fig 7 illustrates a simple half-wave phase sensitive detector. A practical phase-sensitive detector may differ considerably from that illustrated, but Fig. 7 has been chosen for ease of illustration and explanation.

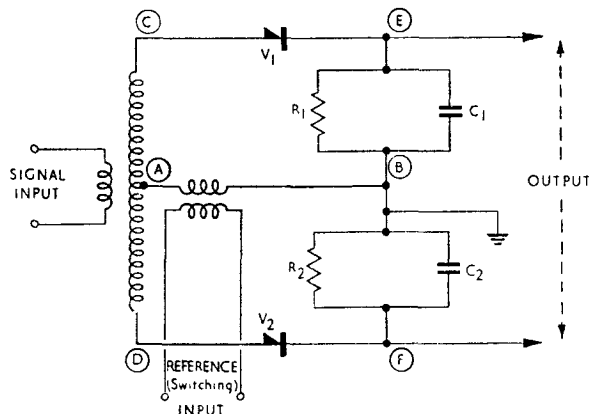


FIG. 7. SIMPLE PHASE-SENSITIVE DETECTOR

If we consider only the reference (switching) waveform, we can see that on the half-cycle when A is positive with respect to B, both diodes are forward-biased and conduct. If there is *no signal* input, both diodes conduct equally through the equal loads R_1 and R_2 and the voltages developed across EB and FB are equal and opposite; they therefore cancel so that there is no output across EF. On the other half-cycle of the reference input, both diodes are cut off and again there is no output.

Let us now apply a signal and assume that C goes positive with respect to D. On the half cycle of reference voltage when both diodes are cut off,

the signal voltage is insufficient to overcome the reverse bias on the diodes and the output across EF remains at zero. When, however, the diodes are forward-biased by the reference voltage and, at the same time, the signal causes C to become positive with respect to D, then V_1 conducts more heavily than V_2 so that E becomes positive with respect to F. If the signal polarity is reversed so that C is negative with respect to D during the time that the diodes are forward-biased by the reference voltage, then the output terminal E becomes negative with respect to F. Thus when C is positive at the same time as A (i.e. signal and reference voltages *in phase*) we get a *positive-going* output; when C is negative at the time A is positive (i.e. signal and reference voltages *in anti-phase*) we get a *negative-going* output.

Now let us examine the circuit with various phase relationships between reference and signal inputs. The two inputs are assumed to be 'in phase' in this example when the signal causes C to be at its maximum positive value at the same instant as the reference voltage causes A to be at its maximum positive value. The results of different phase relationships between input signal and reference voltage are then as shown in Fig. 8.

In *a* the signal is in phase with the reference waveform and the output has its maximum positive value. In *b* the signal is lagging 45° on the reference voltage and, although the output is still positive, its magnitude is less than in *a*. In *c* the signal is lagging 90° on the reference voltage, the shaded areas above and below the zero line are equal, and the output is zero. In *d* the signal is lagging the reference voltage by 135° , the shaded area below the line is now greater than that above it, and the output becomes negative. When the signal is in anti-phase with the reference voltage as in *e*, the output has its maximum negative value.

Thus the polarity and magnitude of the output voltage is a measure of the phase relationship between the signal and the reference voltage. If the phase relationship is *constant* (i.e. when signal and reference inputs are at the *same frequency*) the output is *constant*. But if the phase relationship *varies* (i.e. when the signal and reference inputs *differ* in frequency) then the output will *vary* in polarity and magnitude. In fact, in the latter case, the variation in output voltage takes the form of a cosine wave if plotted against the phase difference between signal and reference voltages (Fig. 9).

Applying this principle to the Doppler radar, the reference voltage is at the frequency f_{IF} derived from the transmitter, and the signal input is the Doppler-shifted echo at $f_{IF} \pm f_D$ from the receiver. Because of the slight *difference* in frequency between the two inputs due to f_D , there is a *continuously changing* phase difference between the two. The output from the phase-sensitive detector then takes the form shown in Fig. 9 and is the detected Doppler shift.

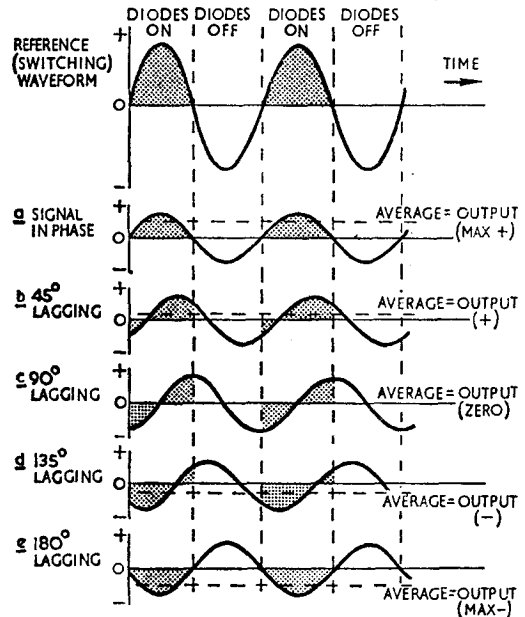


FIG. 8. EFFECT OF VARIOUS PHASE RELATIONSHIPS BETWEEN SIGNAL AND REFERENCE VOLTAGES

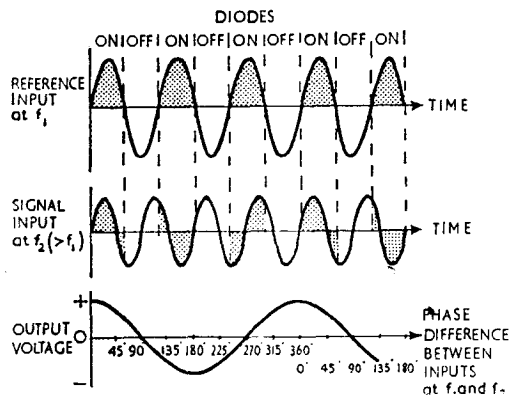
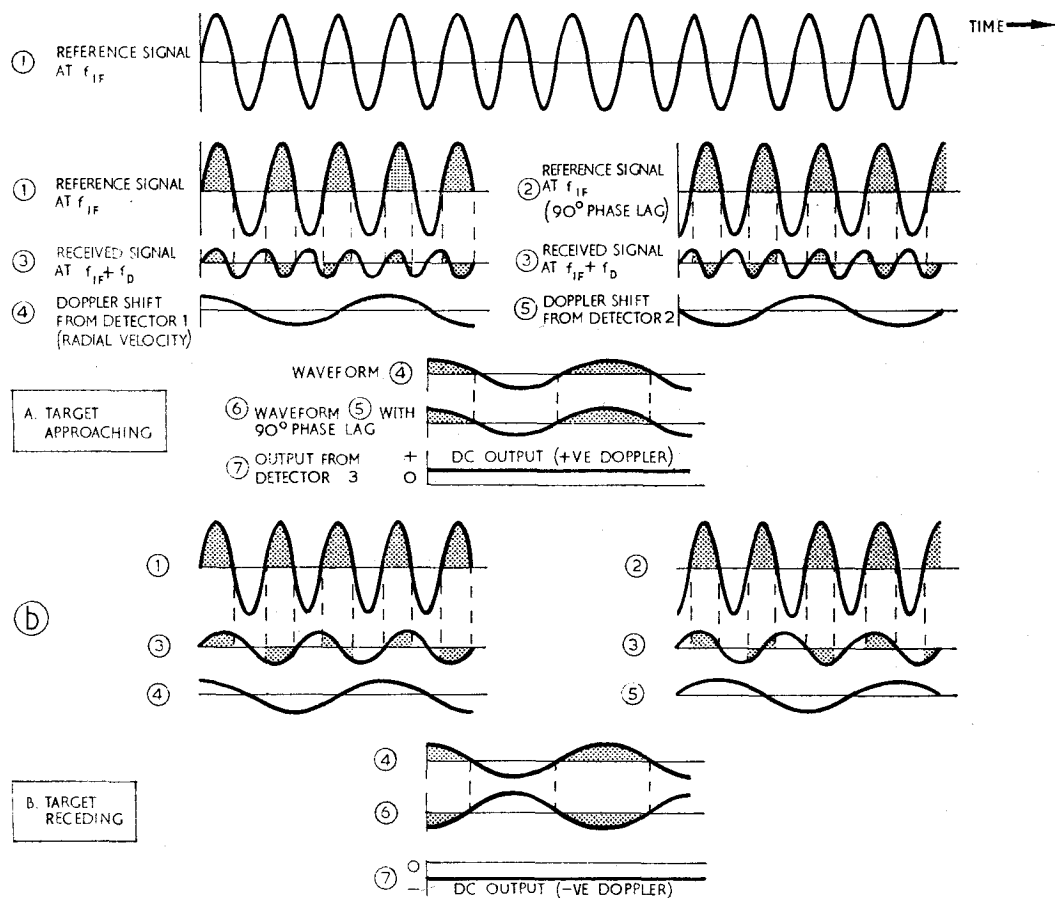
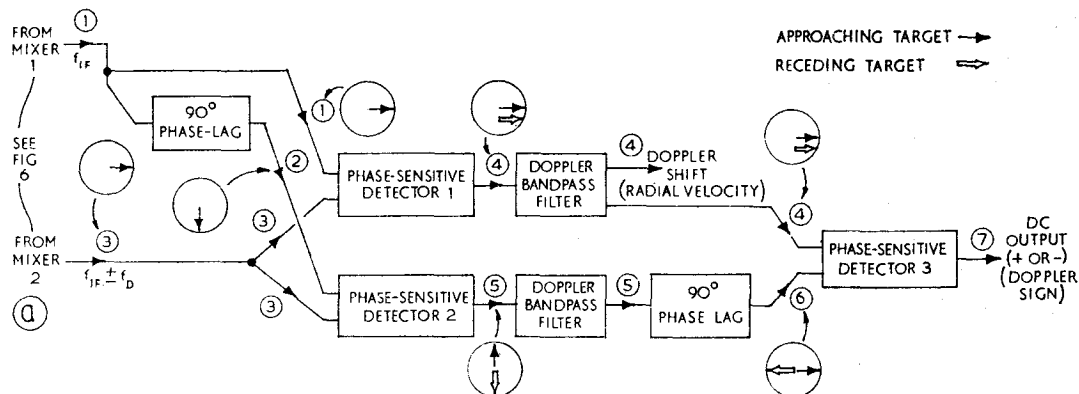


FIG. 9. VARIATION OF OUTPUT WITH VARIATION OF PHASE DIFFERENCE BETWEEN INPUTS

Determination of Sign of Relative Velocity

We saw earlier that if a bank of Doppler filters were used in the i.f. stage of a c.w. receiver (Fig. 5), it would be fairly easy to determine whether a target was approaching or moving away from a ground radar. In this way the direction of motion of the target is resolved and we have *unambiguous Doppler*.



NOTE. SHADED AREAS OF WAVEFORMS INDICATE CONDUCTION OF PHASE-SENSITIVE DETECTORS

FIG. 10. DETERMINING THE "SIGN" OF THE DOPPLER SHIFT

It is not always desirable to use a filter bank in the i.f. stage. Such an arrangement requires a large number of filters *either side* of the centre i.f.; and the accuracy of the system is not sufficiently high for some applications. A filter bank or a velocity gate to measure the relative velocity *after* the Doppler signal has been extracted from the i.f. is usually preferred (see later). In such cases, other means must be used to determine the *sign* of the Doppler shift.

One method of resolving the ambiguity of the Doppler shift is illustrated in Fig. 10. This arrangement uses a combination of phase-sensitive detectors and circuits which introduce a 90° phase lag. The action of the circuit may be explained with reference to the waveforms of Fig. 10. The corresponding phase vectors are also indicated on the block diagram.

The reference signal at f_{IF} (waveform 1) is derived from the transmitter output as explained earlier. This is applied to phase-sensitive detector 1, along with the Doppler-shifted signal at $f_{IF} \pm f_D$ (waveform 3) derived from the receiver. The combination of these two inputs in detector 1 produces the Doppler shift (waveform 4) as explained in Fig. 9. The Doppler frequency is then applied through a bandpass filter to remove unwanted components, and the Doppler output from the filter may be measured to give the radial velocity of the target.

The reference signal (waveform 1) is also applied to a 90° phase-lag circuit, the output of which is waveform 2. Waveform 2 is applied to detector 2, along with the Doppler-shifted received signal (waveform 3). The output of detector 2 also represents the Doppler shift (waveform 5) but it is in *phase quadrature* to the output of detector 1 (waveform 4).

Waveform 5 is applied through a Doppler bandpass filter to a 90° phase-lag circuit, the output of which is waveform 6. The relative phase of waveform 4 and 6 are then compared in detector 3, the output of which is a *d.c. voltage of polarity* depending upon the direction of motion of the target. If the target is *approaching* the radar, waveforms 4 and 6 are *in phase*, and this produces a *positive* d.c. output voltage from detector 3, indicating a positive Doppler shift. If the target is *moving away* from the radar, waveforms 4 and 6 are *in anti-phase*, and this produces a *negative* d.c. output voltage from detector 3, indicating a negative Doppler shift.

Another method of producing an unambiguous Doppler output is illustrated in Fig. 11. The arrangement is the same as that of Fig. 10 up to and including the Doppler bandpass filters. The associated vectors assist in the explanation of the circuit action.

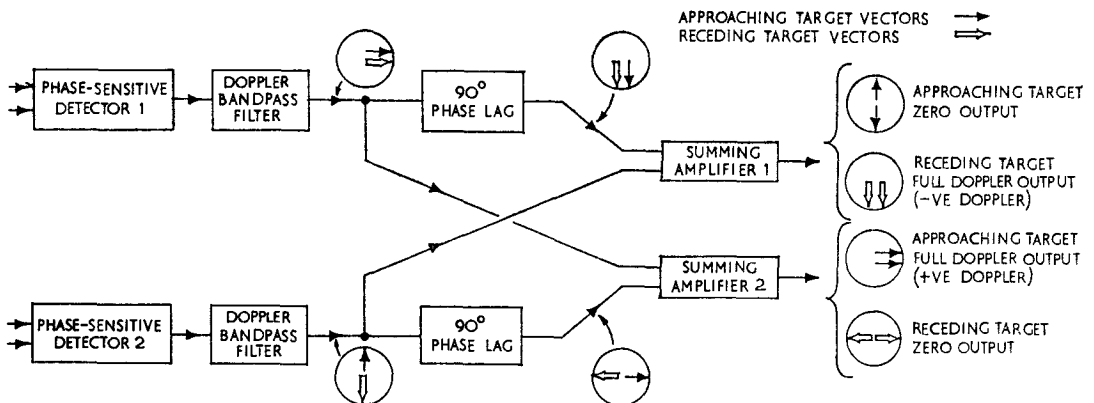


FIG. 11. UNAMBIGUOUS DOPPLER EXTRACTION CIRCUIT

The output of detector 1 is subjected to a 90° phase delay and applied to summing amplifier 1, along with the undelayed output of detector 2. For *approaching* targets these two signals are *in anti-phase* and the output from summing amplifier 1 is *zero*. For *receding* targets the two signals are *in phase* and the output from summing amplifier 1 is a Doppler shift signal proportional in amplitude to the *sum* of the two inputs. Thus, summing amplifier 1 produces an output only for *receding* targets.

In a similar way it may be seen that summing amplifier 2 produces an output only for *approaching* targets. Hence, a Doppler output from summing amplifier 1 indicates a *negative* Doppler; if the Doppler output is from summing amplifier 2, a *positive* Doppler is indicated. The circuit arrangement is such that the active channel produces an appropriate indication. The Doppler output from either channel is applied to frequency-measuring circuits for indication of radial velocity.

There are other methods of determining the sign of the Doppler shift. For example, we could dispense with part of the circuit of Fig. 11 and apply the Doppler outputs of the two bandpass filters to a *two-phase induction motor* (see p 567 of Part 1B). Since the output from filter 1 is always 90° leading or lagging that from filter 2, the application of these two inputs to the motor causes the motor to rotate. The direction of rotation depends upon the direction of motion of the target. The Doppler sign may thus be indicated.

Measurement of Radial Velocity

So far we have seen how the Doppler shift is extracted and how its sign is determined. The next step is the measurement of the Doppler shift as a means of indicating the radial velocity of the target. We have already seen that it is possible to measure the Doppler shift by using a bank of Doppler filters in the i.f. stage of the receiver before the detector. It is also possible to use a bank of narrow-band filters after the Doppler-extraction circuits. Because of the action of the Doppler-extraction circuits only the band of Doppler frequencies either above or below the i.f. is passed. Thus only *half* the number of filters are needed compared with that used by the i.f. filter bank. In addition, the passband of each filter can be made much narrower within the Doppler range of frequencies than when the filters are at the i.f. The accuracy of Doppler measurement is thereby improved.

Fig. 12 shows an arrangement in which a bank of filters covering the required Doppler frequency range is inserted after the Doppler-extraction circuit. Each filter is centred on a slightly

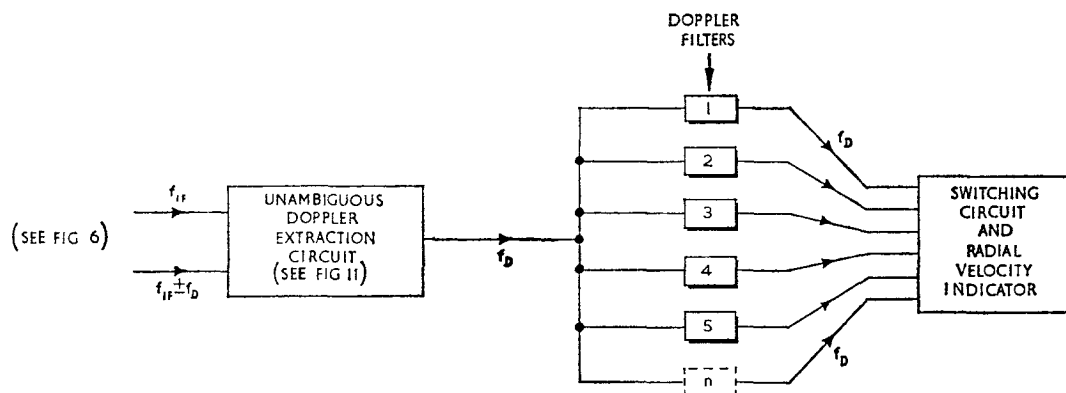


FIG. 12. USE OF DOPPLER FILTER BANK AFTER DOPPLER EXTRACTION CIRCUIT

different frequency and passes only a very narrow band of frequencies, the passband being determined by the required accuracy (125 c/s passband is typical). The filters can be switched in sequence, and the arrangement is such that the switching sequence stops at the filter which shows an output. This filter then provides an indication of radial velocity.

Measurement of the radial velocity can also be obtained from the circuit arrangement known as a *velocity tracking gate* (Fig. 13). The Doppler frequency output from the Doppler-extraction circuit is applied to a mixer where it combines with the output of a variable frequency oscillator. The frequency of the oscillator is controlled by a modulator circuit (e.g. a reactance

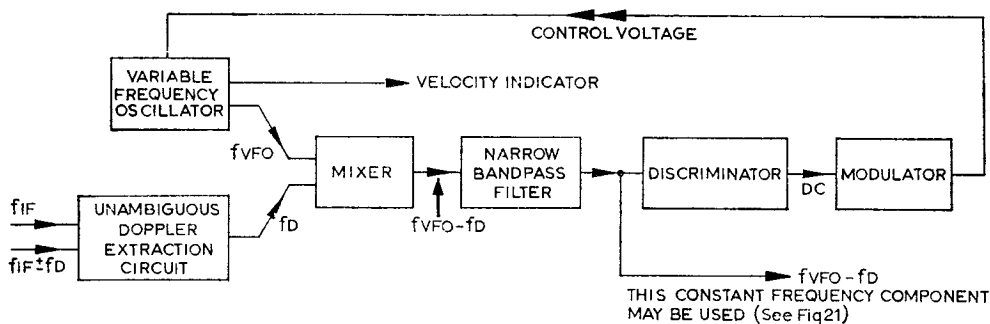


FIG. 13. VELOCITY TRACKING GATE

valve or a varactor diode) and the circuit is so designed that the oscillator operates over a frequency range such that $f_{VFO} - f_D$ is a *constant frequency* for the expected range of Doppler shifts. The output of the mixer should therefore be a constant frequency and this is passed to a narrow-band filter. The filter output is applied to a discriminator which, in turn, drives the modulator and so controls the frequency of the oscillator.

Let us consider an example. Suppose the narrow-band filter is centred on 100 kc/s. Then at all times, $f_{VFO} - f_D$ should equal 100 kc/s. If the Doppler frequency output of the Doppler-extraction circuit is at 5 kc/s, f_{VFO} must be at 105 kc/s to keep the filter output at 100 kc/s. If now the Doppler increases to 6 kc/s, the filter output falls instantaneously to 99 kc/s. The discriminator and modulator then become operative to adjust the frequency of the oscillator to 106 kc/s. Thus, the oscillator frequency *varies in step* with the Doppler frequency and is therefore a measure of the radial velocity. In a null-measuring arrangement like this, the accuracy of measurement is high. Note that the velocity tracking gate may be used to track a selected target in terms of its radial velocity. If the selected Doppler frequency varies due to a change in the relative velocity of the target, the velocity gate automatically 'follows' this variation in frequency.

Noise

The causes of receiver noise and the steps taken to reduce its effect and improve the receiver noise factor are dealt with elsewhere in this book (see p 397). However, in c.w. radar, noise is particularly troublesome. The continuous transmission results in continuous returns from all reflecting objects and those from nearby objects may be very strong. In addition, a portion of the transmitted signal is *intentionally* applied to the receiver as a reference against which the Doppler-shifted echo is compared. Any noise in the *transmitted* signal therefore appears in the receiver and reduces the receiver sensitivity. In a c.w. radar it is essential to reduce *transmitter noise* to a minimum. Transmitter noise results mainly from random variations in the amplitude and frequency of the transmitted signal, caused by variations in power supplies and temperature, and by mechanical vibration. Such variations are kept to a minimum by rigid stabilization of power supplies, by the use of temperature-control equipment, and by the reduction of all forms of vibration. This reduces the noise content of the transmitted signal to a low level. Where further reduction is required for optimum performance of the c.w. radar, *noise control circuits* may be inserted. These circuits effectively extract the noise components from the transmitted signal and feed them back in anti-phase into the transmitter circuits (negative feedback loop) to cancel the noise in the transmitter output.

Clutter

The received signal consists of clutter (returned mainly from fixed objects) and the required moving target signal which is frequency-shifted from the transmitter frequency by the Doppler shift. Clutter signals returned from objects near the transmitter will be at a very high level and

may be sufficient to saturate the signal crystal mixer so that it becomes insensitive to weak signals returned from long-range targets. It is therefore necessary to reduce the level of clutter before it reaches the crystal mixer.

The reflection from a stationary object is at the same frequency as that of the transmitted signal, i.e. clutter signals from stationary objects have zero Doppler shift. In addition, slow-moving ground objects (e.g. trees blowing in the wind) have a very low Doppler shift and also produce clutter, so that the composite clutter signal has a *small finite bandwidth*. Most of the clutter signal lies within about ± 1 kc/s of the transmitter carrier frequency. Thus, in removing most of the clutter signal to obtain a low clutter level we also remove some of the lower Doppler shift frequencies. These lower Doppler shifts correspond to very low radial velocities (about 30 m.p.h. for an X-band radar) and are only important for airborne targets when the target is approaching the crossing point.

One method of reducing clutter is shown in schematic outline in Fig. 14. The outputs of the phase-sensitive detectors contain the Doppler frequency components. For an X-band

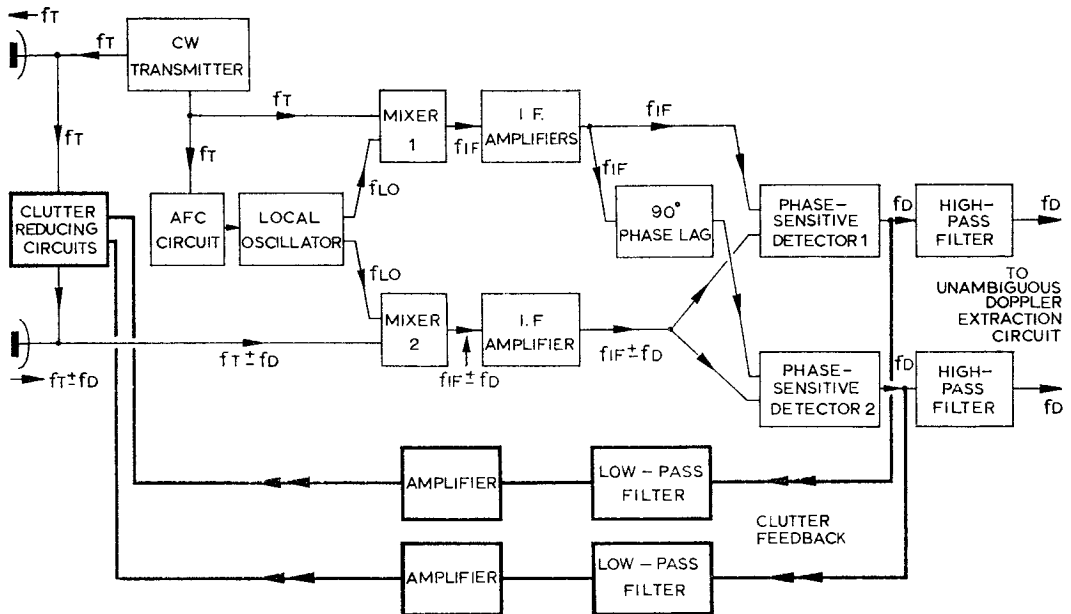


FIG. 14. SCHEMATIC ARRANGEMENT FOR REDUCTION OF CLUTTER

radar operating against high-speed airborne targets, the expected range of Doppler frequencies may lie within the band 0-50 kc/s. The high-pass filters shown have a low-frequency cut-off at about 1 kc/s. The result is that the majority of the Doppler frequencies (in the band 1-50 kc/s) are applied through the high-pass filters to the unambiguous Doppler extraction circuit. Most of the clutter, on the other hand, (in the band 0-1 kc/s) is applied through the low-pass filters. The clutter signals are then amplified and applied as a form of negative feedback to the clutter reducing circuits. There they are impressed on a portion of the transmitter carrier frequency and used to cancel the clutter signals in the input to the receiver.

Block Diagram of CW Doppler Radar System

To consolidate what we have learnt so far about c.w. radar, a schematic diagram of a simple system is illustrated in Fig. 15. This block diagram uses all the principles discussed in the previous pages and should be self-explanatory.

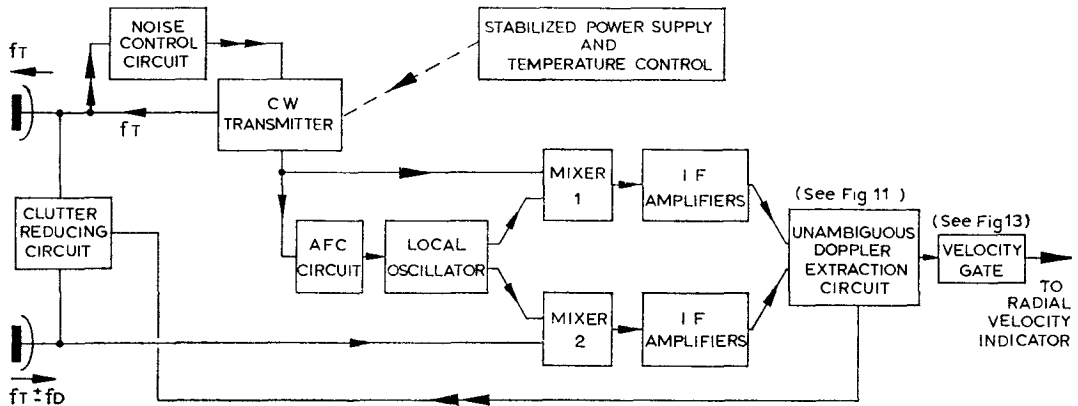


FIG. 15. BLOCK DIAGRAM OF CW RADAR SYSTEM

Advantages and Limitations of CW Radar

In common with pulsed radar, c.w. Doppler radar can be used to detect a target and to find the bearing and elevation of the target (by noting the angular position of the radar scanner). However, when we compare c.w. radar with pulsed radar in more detail we notice that c.w. radar has certain advantages and certain limitations.

1. **Bandwidth.** In a pulsed radar, the required bandwidth is inversely proportional to the pulse duration of the transmitted signal. For good range discrimination short-duration pulses are necessary, and the required bandwidth is then very wide, e.g. a $1\mu\text{s}$ pulse requires a bandwidth of about 2 Mc/s to preserve the pulse shape in the receiver. In a c.w. radar, the Doppler-shifted signal is within a relatively narrow band a few kc/s either side of the transmitted frequency. For the normal expected values of relative velocities a typical bandwidth may be 100 kc/s for an X-band radar; very often it is less than this. Since noise is proportional to bandwidth, the narrow band c.w. receiver has a higher signal-to-noise ratio than that of the wideband pulsed radar receiver; the c.w. Doppler receiver sensitivity is, therefore, better. Wideband noise jamming, such as that produced by a carcinotron (see p 441), also has much less effect on the narrow-band c.w. receiver. A carcinotron jammer producing a noise output of 4 watts per megacycle delivers 8 watts of noise in a 2 Mc/s bandwidth (pulsed radar) compared with 0.4 watts in a 100 kc/s bandwidth (c.w. radar).

2. **Power requirements.** The maximum radar range in both pulsed and c.w. systems is proportional to $\sqrt[4]{P_T/B}$ where P_T is the radiated power and B is the bandwidth. In pulsed radar, B is very large so that P_T must be large also to provide adequate range; peak powers of the order of megawatts are common. In c.w. radar, for a comparable range, P_T can be much smaller since B is also relatively small. Typical figures for pulsed and c.w. radars of comparable range are given in Table 1.

	Pulsed Radar	CW Doppler Radar
Radiated Power (Peak)	2 MW	2 kW
Pulse Duration	$1\mu\text{s}$	—
PRF	1,000 p.p.s.	—
Bandwidth	2 Mc/s	100 kc/s

TABLE 1. POWER AND BANDWIDTH REQUIREMENTS

Although in the example above, the *peak* power radiated by the pulsed radar is 1,000 times greater than that radiated by the c.w. radar, the *average* power is the same for both. For the c.w. radar, the duty factor is unity (transmitter operating continuously) so that peak power equals average power. For the pulsed radar:

$$\begin{aligned}\text{Average power} &= \text{Peak power} \times \text{Pulse duration} \times \text{PRF} \\ &= 2 \times 10^6 \times 1 \times 10^{-6} \times 10^3\end{aligned}$$

$$\text{Average power} = 2 \text{ kW} = \text{CW radar power.}$$

The c.w. radar has the advantage that it is not subject to the very high peak powers of pulsed radar. There is, therefore, less danger of electrical breakdown in the c.w. radar. In addition, no high-power modulator stage is required.

3. **Minimum range.** A pulsed radar has a certain minimum range whose value depends mainly upon the pulse duration of the transmitter. At ranges less than the minimum range, the pulsed radar is 'blind' because the echo is received back at the radar whilst the transmitter is still firing. In c.w. Doppler radar, the transmitter and receiver both operate continuously so that, in theory, the c.w. system can operate down to zero range. In practice, very small minimum ranges are possible.

4. **Discrimination between moving and stationary objects.** In a conventional pulsed radar there is no easy way of distinguishing between moving and stationary objects. The result is that wanted moving targets are often obscured by the returns from unwanted stationary objects (clutter), especially at short ranges and at low angles of elevation. In a c.w. system, the received signal from a moving target is shifted in frequency from the transmitted frequency by the Doppler shift. Reflections from stationary objects produce no Doppler shift. It is therefore possible for a c.w. radar to distinguish between moving and stationary objects by measuring any Doppler shift. Ground clutter can therefore be rejected to allow targets to be seen down to low ranges. Slow-moving clutter signals, such as those from clouds, can also be eliminated by selecting only the higher Doppler frequency shifts (see p 427 on 'clutter').

Targets are normally tracked in c.w. radar by tracking the Doppler shift produced by the target's radial velocity. The *rate of change* of Doppler shift is normally slow enough to allow a velocity tracking gate to follow the target. If, however, the target drops *window* (see p 440), the rate of change of the Doppler return from the window will be too rapid for the velocity tracking gate to follow. Therefore, c.w. tracking radars are much less prone to the effects of window jamming than are pulsed radars.

5. **Inability to determine range.** The greatest limitation of the pure c.w. Doppler radar is its inability to determine the *range* of a target. To measure range, some form of 'mark' must be applied to the transmission to indicate the instant of transmission. The elapsed time between making this mark and the return of the received echo then gives a measure of the range to the target. This is done in pulsed radar, where the mark is the transmitted pulse. The sharper the mark (i.e. the narrower the pulse in pulsed radar), the greater is the accuracy of the elapsed-time measurement, and hence range measurement. A widely-used method of inserting a distinguishing timing mark in a c.w. system in order to measure range is to *frequency-modulate* the r.f. carrier.

Method of Ranging in CW Radar

We have seen that a pure c.w. radar can be used to determine the bearing and elevation of a target, and also its radial velocity and 'sign'. In addition, the target can be tracked in terms of its radial velocity. Where *range information* is also a requirement, the transmitter may be *frequency-modulated* at a low modulation frequency as shown in Fig 16. Fig 16a illustrates the waveform of the modulating signal and Fig. 16b shows the waveform of the resulting f.m.c.w. transmitter output. The frequency of the transmitter output varies *at a rate* determined by the *frequency* of the modulating signal, whilst the *amount* of frequency deviation depends

upon the *amplitude* of the modulating signal. Very often it is easier to represent the frequency-modulated transmitter output in the form shown in Fig. 16c; this graph shows the *frequency* of the transmitter plotted against time, and the frequency deviation is clearly seen.

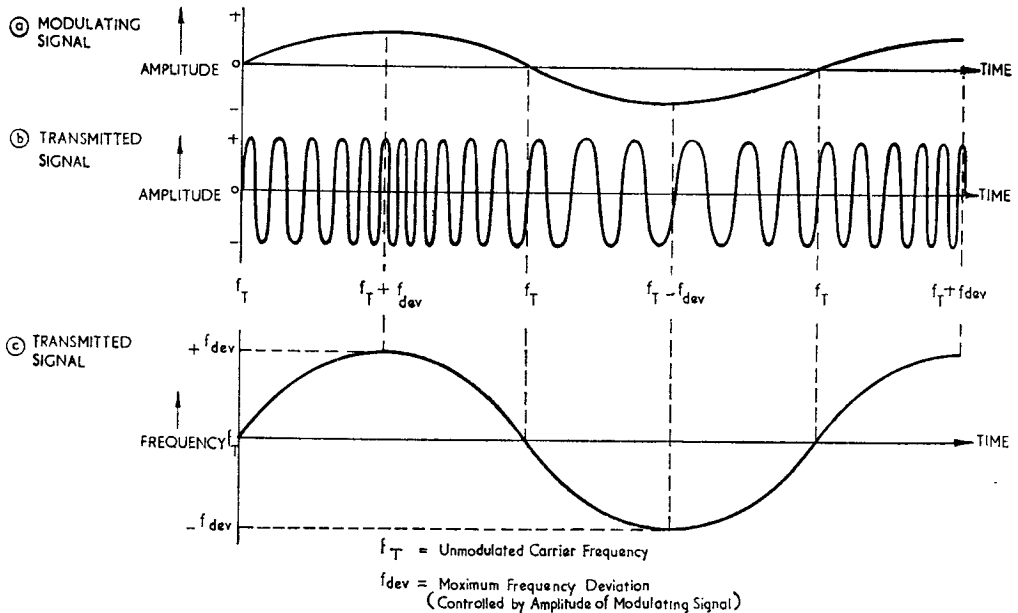


FIG. 16. FREQUENCY-MODULATION WAVEFORMS

For the moment, to avoid confusion, we shall ignore the Doppler shift in received frequency produced by a moving target and concentrate on the measurement of range of a *stationary* target (zero Doppler). Provided that the modulation frequency is chosen to suit the range to be measured, the range can be determined by comparing the *instantaneous phase* of the received signal modulation with that of the transmitted signal modulation. Due to the time taken for a signal to return from a target, the phase of the return modulation will *lag* that of the transmitted signal modulation, i.e. the received signal modulation has the phase that the transmitted signal modulation had *some time earlier*. If a portion of the transmitted signal is now mixed with the returned echo signal, then the output of the mixer will contain a component which varies in frequency at the same rate as the original signals, but with a much smaller frequency deviation. The resultant wave is called the *reduced deviation signal* (Fig. 17).

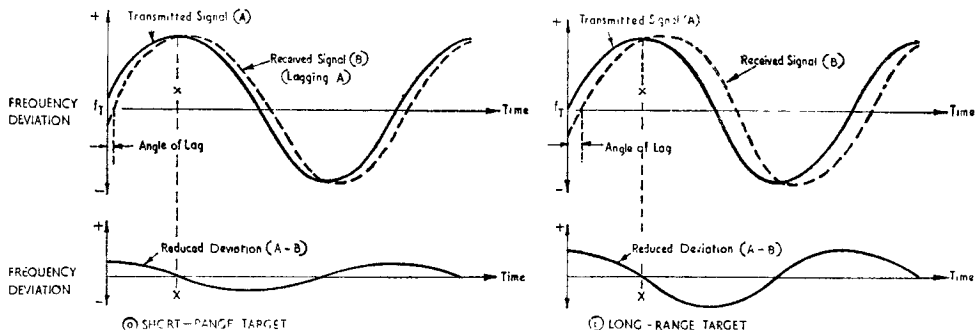


FIG. 17. REDUCED DEVIATION AND ITS VARIATION WITH RANGE

The phase difference between the modulation of the received signal and that of the transmitted signal is proportional to the elapsed time between transmission and reception and is, therefore, proportional to target range. Fig. 17a shows a small angle of lag between received and transmitted signal modulations, and corresponds to a close-range target. Because of the small phase difference, the reduced deviation is also small. The conditions for a longer-range target are illustrated in Fig. 17b. The delay between received and transmitted signals is now greater and the increased phase difference produces a larger reduced deviation. For small delays the *magnitude* of the reduced deviation is *proportional to range*.

Frequency of Modulating Signal

On p 35 we showed that, in a pulsed radar, the p.r.f. is selected according to the maximum required range. For long-range detection, a low value of p.r.f. is required to ensure that the echo from the most distant required target has returned before the transmitter fires again. There is a similar requirement for the *frequency* of the *modulating signal* used for ranging in c.w. radar. In comparing the phase of the received signal modulation with that of the transmitted signal, only the first 90° of the modulating cycle can be used without introducing ambiguities. A circuit cannot easily distinguish between, say, 45° phase lag and 135° phase lag; Fig 17 shows that for equal phase differences *either side* of point X the reduced deviation has the same magnitude. Thus, for unambiguous range measurement, since only the first quarter of the cycle can be used, the period occupied by one cycle of the modulating signal must be *at least four times* as long as the time taken for a signal to return from the most distant required target. For long-range detection of targets *low* modulation frequencies are, therefore, required; a typical value is 30 c/s.

To avoid an unduly large reduced deviation, only about the first 20° of the modulating cycle are used for ranging purposes (Fig. 17b shows an angle of lag of about 36°). For a modulation frequency of 30 c/s, one cycle has a period of $1/30$ second. For a phase difference of 18° ($1/20$ of a cycle) between received and transmitted signals, the time delay is then $1/600$ second. Since e.m. energy travels at a speed of 186,000 miles per second it can be seen that in $1/600$ second the energy has travelled 310 miles to the target and back. This gives a *target range* of 155 miles. Because of the relatively small phase differences between received and transmitted signals, and the resulting small reduced deviation, the *accuracy of ranging* in a c.w. radar system of this type is not high. However, approximate values of range are sufficient for most applications, because tracking of the target is carried out in terms of radial velocity.

Combined Range and Doppler Signals

In the discussion so far on ranging, no mention has been made of the effect of Doppler. From previous notes we know that when a target is in motion the received signal is shifted in frequency either above or below the transmitted frequency by an amount equal to the Doppler shift f_D . If the transmitter is frequency-modulated to provide ranging, the *whole frequency spectrum* of the frequency-modulated received signal is shifted either above or below the transmitted frequency depending upon whether the target is approaching or moving away from the radar. This is illustrated in Fig. 18.

In Fig. 18a the target is approaching the radar and the received signal at $f_T + f_D$ is *higher* than the transmitted signal by the Doppler shift f_D . Superimposed on both the transmitted and received signals we have the ranging modulation f_{dev} . The received ranging modulation is delayed in phase relative to the transmitted modulation by an amount proportional to the range to the target. If now we plot the graph of the instantaneous difference in phase between transmitted and received signals, the reduced deviation waveform is produced as previously described. However, it is now impressed on the Doppler frequency f_D . The waveforms for a *receding* target can be worked out in the same way and are as shown in Fig. 18b.

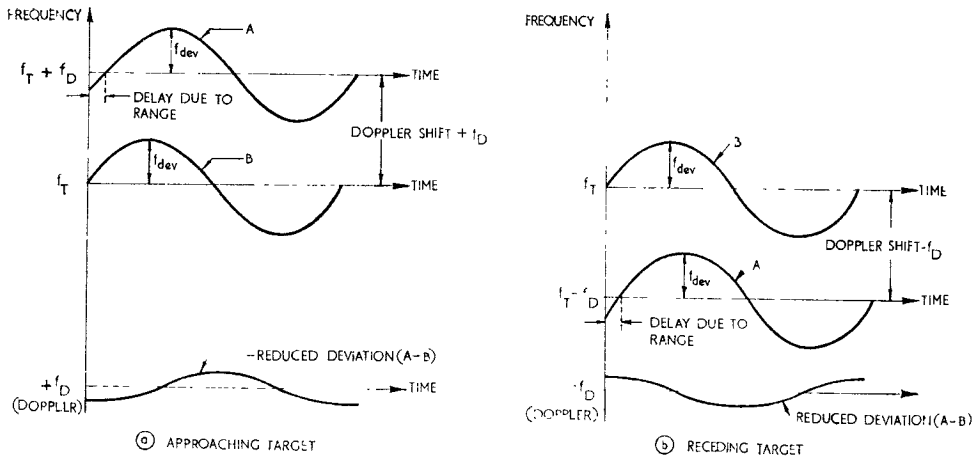


FIG. 18. EFFECT OF DOPPLER ON RANGE MODULATION SIGNALS

We shall see later that the extracted Doppler signals in the receiver *retain* the ranging modulation, so that measurement of range can be carried out on the signal at the selected Doppler frequency (for the selected target).

Fig. 19 shows the basic block diagram of a c.w. radar which is used to measure both range and radial velocity. The associated waveforms are illustrated in Fig. 20.

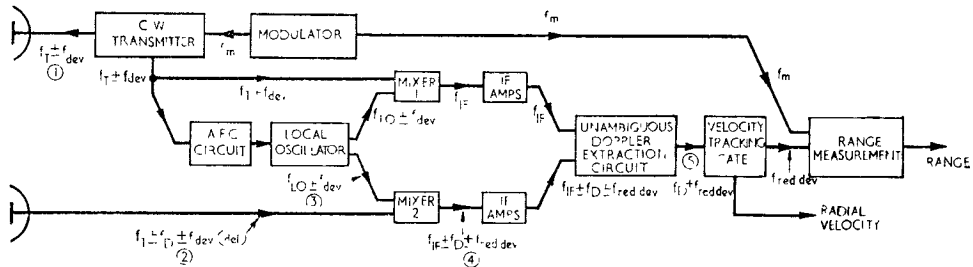


FIG. 19. CW RADAR SYSTEM FOR MEASUREMENT OF RANGE AND RADIAL VELOCITY

The transmitter is frequency-modulated by the modulating signal f_m to produce an r.f. output at $f_T \pm f_{dev}$, where f_{dev} is the frequency deviation used for ranging (waveform 1). The received signal from a moving target at a given range is then $f_T \pm f_D \pm f_{dev(del)}$, where f_D is the Doppler shift and $f_{dev(del)}$ is the ranging modulation, delayed in phase because of the range to the target (waveform 2). In Fig. 20, waveform 2 represents a *receding* target (negative Doppler $-f_D$).

The frequency of the local oscillator is being controlled by an a.f.c. circuit whose input is the transmitter reference signal at $f_T \pm f_{dev}$. The a.f.c. circuit responds to the relatively slow rate of change of the deviation signal and so varies the local oscillator frequency to give an output at $f_{LO} \pm f_{dev}$ (waveform 3). The local oscillator deviation is *in phase* with the transmitter deviation so that when the local oscillator output and the received signal (waveforms 2 and 3) are combined in the receiver mixer, the output from the mixer is at $f_{IF} \pm f_D \pm f_{red dev}$, where $f_{red dev}$ is the reduced deviation (waveform 4). Again in Fig. 20, waveform 4 is for a *receding* target ($-f_D$).

The unambiguous Doppler extraction circuit extracts the Doppler and reduced deviation components $f_D \pm f_{red dev}$ (waveform 5), and also determines the *sign* of the Doppler (see p 453). The Doppler and reduced deviation components are then applied to the velocity tracking gate,

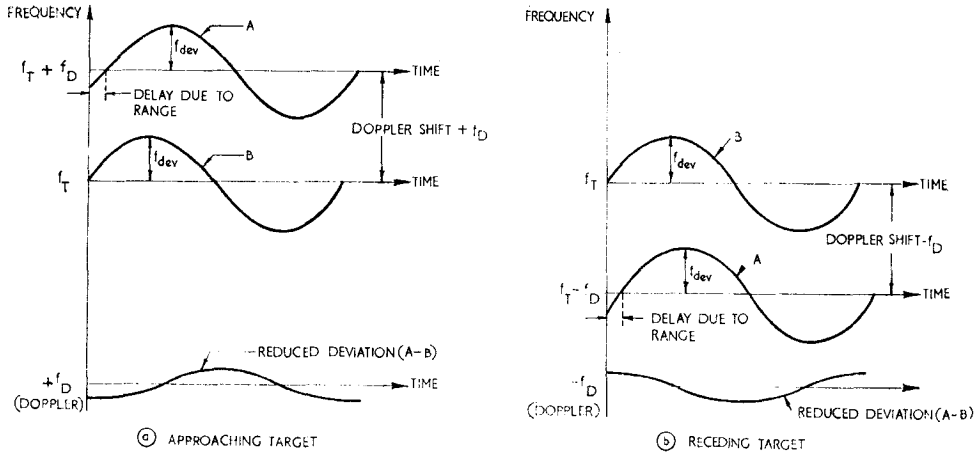


FIG. 18. EFFECT OF DOPPLER ON RANGE MODULATION SIGNALS

We shall see later that the extracted Doppler signals in the receiver *retain* the ranging modulation, so that measurement of range can be carried out on the signal at the selected Doppler frequency (for the selected target).

Fig. 19 shows the basic block diagram of a c.w. radar which is used to measure both range and radial velocity. The associated waveforms are illustrated in Fig. 20.

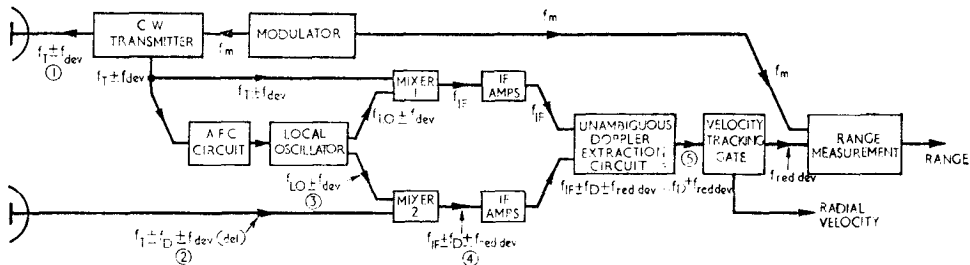


FIG. 19. CW RADAR SYSTEM FOR MEASUREMENT OF RANGE AND RADIAL VELOCITY

The transmitter is frequency-modulated by the modulating signal f_m to produce an r.f. output at $f_T \pm f_{dev}$, where f_{dev} is the frequency deviation used for ranging (waveform 1). The received signal from a moving target at a given range is then $f_T \pm f_D \pm f_{dev(del)}$, where f_D is the Doppler shift and $f_{dev(del)}$ is the ranging modulation, delayed in phase because of the range to the target (waveform 2). In Fig. 20, waveform 2 represents a *receding* target (negative Doppler $-f_D$).

The frequency of the local oscillator is being controlled by an a.f.c. circuit whose input is the transmitter reference signal at $f_T \pm f_{dev}$. The a.f.c. circuit responds to the relatively slow rate of change of the deviation signal and so varies the local oscillator frequency to give an output at $f_{LO} \pm f_{dev}$ (waveform 3). The local oscillator deviation is *in phase* with the transmitter deviation so that when the local oscillator output and the received signal (waveforms 2 and 3) are combined in the receiver mixer, the output from the mixer is at $f_{IF} \pm f_D \pm f_{red dev}$, where $f_{red dev}$ is the reduced deviation (waveform 4). Again in Fig. 20, waveform 4 is for a *receding* target ($-f_D$).

The unambiguous Doppler extraction circuit extracts the Doppler and reduced deviation components $f_D \pm f_{red dev}$ (waveform 5), and also determines the *sign* of the Doppler (see p 453). The Doppler and reduced deviation components are then applied to the velocity tracking gate,

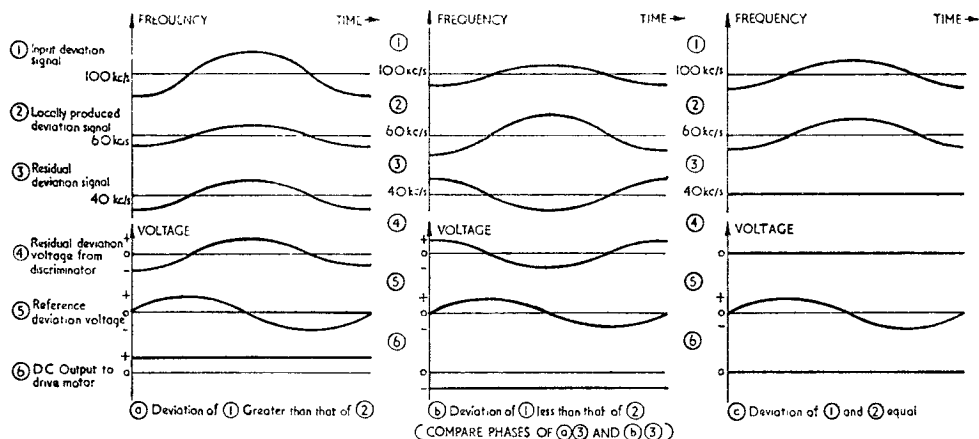


FIG. 22. ACTION OF RANGE-MEASUREMENT CIRCUIT

This *residual deviation signal* is illustrated by waveform 3. The discriminator is centred on 40 kc/s so that if the input to the discriminator is deviating in frequency about 40 kc/s, a voltage output is produced (waveform 4). This voltage is applied to a phase-sensitive detector, together with a reference signal f_M from the transmitter modulator (waveform 5). The output from the phase-sensitive detector is then a d.c. voltage, either positive or negative in polarity (waveform 6) depending upon the relative magnitudes of waveforms 1 and 2. This d.c. output is used to drive a motor, the output shaft of which drives a synchro (for indication of range) and also the wiper of a range potentiometer. The potentiometer controls the amplitude of the modulating signal f_M applied to the local oscillator and, in this way, controls the frequency deviation of waveform 2.

The arrangement is such that the frequency deviation of waveform 2 is brought into equality with that of waveform 1 (Fig. 22c). When the two deviations are equal, there is no output from the discriminator and the motor stops. The position at which the motor shaft and potentiometer wiper stop is a measure of range to the target, and a synchro driven by the motor shaft can be calibrated in terms of range.

If the target range now increases, the reduced deviation signal increases in magnitude (see Fig. 17) and the circuit again becomes operative to increase the frequency deviation of the local oscillator signal. When the two are balanced once more, the motor stops and the increased range is indicated. The range measurement circuit therefore tracks in range that target which has been selected by the velocity gate.

Other ways of measuring the range and relative velocity of a target have been used in practice. Most of these methods use some form of frequency modulation as the ranging component, and all of them have the disadvantage that range accuracy is sacrificed for accuracy in relative velocity measurement. However, as we have seen, accurate ranging is not usually a requirement for ground radar systems.

FMCW RADAR ALTIMETER

Introduction

Aircraft altimeters are used to give a continuous indication of the *height* at which an aircraft is flying. What, however, do we understand by 'height'? Do we mean the distance to the ground beneath the aircraft; or the height of the aircraft above sea level? In a mountainous region, the answers are quite different. In fact, modern aircraft are fitted with *two* altimeters: a *radio (radar) altimeter*, which indicates the height of the aircraft above the ground immediately beneath (a terrain-clearance indicator); and a *barometric altimeter*, which indicates height above a reference level (usually sea level). The differences between the two types are illustrated in Fig. 1.

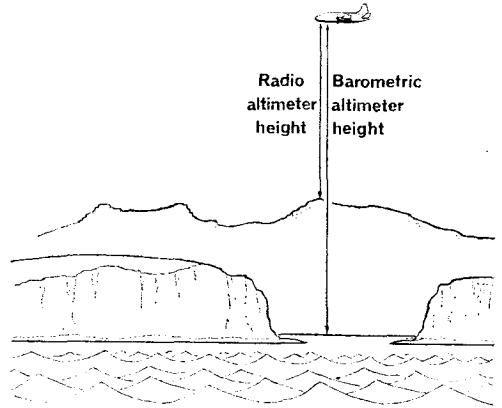


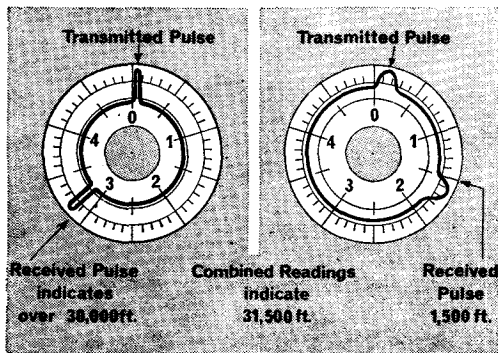
FIG 1. RADIO AND BAROMETRIC ALTIMETERS

Both instruments are essential. An aircraft must be capable of maintaining straight and level flight over any type of terrain at a given height above sea level. The pre-arranged flight plan may call for this; or, to avoid collision in an air corridor, a ground controller at an airfield may require it. In either case, because the height indication of a radio altimeter follows the contours of the ground over which the aircraft is flying, the *barometric altimeter* is essential.

However, the barometric altimeter depends for its operation on *changes* in atmospheric pressure and, since the pressure at sea level may vary in a random manner, the barometric altimeter reading can be in error by several hundreds of feet at any given time. This is not accurate enough for low-level flying. For accurate indication of height above the ground—especially when the aircraft is landing—the *radio altimeter* is used.

Pulse-modulated Radar Altimeter

In a radar altimeter, height is determined by measuring the time delay between the transmission of a signal and its reception back at the aircraft via the ground. The time delay may be measured directly, as in the *pulse-modulated* radar altimeter. Alternatively, it may be measured indirectly by calculating the difference in frequency between that of the transmitted and received signals, as in the *frequency-modulated* c.w. radar altimeter. Both types are used in the RAF.



(a) 0-50,000ft. Display (b) 0-5,000ft. Display

FIG 2. PULSE-MODULATED RADAR ALTIMETER DISPLAY

The principle of operation of the pulse-modulated altimeter follows normal radar technique. The time taken for a short-duration pulse to travel to the ground and back is measured and displayed on a c.r.t. indicator. The timebase is synchronized with the transmitter pulses and the p.r.f. of the system is adjusted for the required height range. One pulse-modulated altimeter used in the RAF has two p.r.f.s, 98.356 kc/s and 9.835 kc/s. The first p.r.f. gives a height range of 0-5,000 feet, and the second 0-50,000 feet. A type J circular display is used. To indicate distance to the ground beneath, both range scales and both timebases are used. The 0-50,000 feet range scale is marked on the inner circle of the c.r.t. indicator, and the 0-5,000 feet range scale on the outer circle. The 0-50,000 feet scale is

used to give an approximate indication of height. The 0-5,000 feet range gives the required accuracy. Fig 2 indicates the use of the display on both range scales.

At heights above 500 feet the pulse-modulated radar altimeter reading is accurate to within ± 100 feet. Errors arise from such factors as variations in the p.r.f., non-linearity of the timebases, and difficulty in estimating the exact position of the leading edge of the pulse on the indicator. Above 500 feet, this order of accuracy is adequate.

This does not apply, however, for heights below 500 feet. Like all pulse-modulated radars, the altimeter has a certain *minimum* range which depends mainly on the pulse duration of the transmitted pulse. During the time the transmitter is firing, the receiver is inoperative. For a $0.25 \mu\text{s}$ pulse, the resulting 'blind' distance is 125 feet. In addition, the receiver takes a finite time to recover so that the ineffective range of a pulse-modulated radar altimeter may be of the order of 400 feet. Thus, although adequate for indication of terrain clearance at altitudes above 500 feet, the pulse-modulated radar altimeter—like the barometric altimeter—cannot be used for low-level flying. For this, the f.m.c.w. radar altimeter is used.

Principle of FMCW Radar Altimeter

We dealt very briefly with the f.m.c.w. altimeter in Section 1 of these notes (see p54). There we showed that the altimeter transmitter beams an f.m.c.w. signal down towards the ground. At the ground some of this energy is reflected and received back at the aircraft. However, during the time taken for the signal to travel to the ground and back, the *transmitter* frequency has changed. We thus have a *difference* in frequency between received and transmitted signals at any given instant of time, and this frequency difference is a measure of the height of the aircraft above the ground. Let us now examine this principle in more detail.

Let us suppose that the frequency of the altimeter transmitter is made to increase linearly with time, as shown by the solid line in the graph of Fig 3a. After a certain time some of the transmitted energy will arrive back at the altimeter receiver via the ground. The time interval depends upon the height h of the aircraft and is given by the usual relationship:

$$\text{Time } t = 2 \frac{h}{c} \text{ where } c = \text{velocity of radio waves.}$$

As always in expressions like this, care must be taken in the choice of units. If h is to be given in feet, c must be given in feet per second, and t in seconds. If h is to be given in metres, c will be in metres per second, and t in seconds.

The frequency of the received signal varies in the manner shown by the dashed line of Fig 3a, and is seen to be the same as that transmitted *some time earlier*. Thus at any given instant there is a difference in frequency between transmitted and received signals and if the transmitted signal is applied directly to a mixer along with the indirect signal reflected from the ground a beat frequency f_B is produced as shown in Fig 3b. The value of the beat frequency depends upon the time interval t in Fig 3 and so gives a measure of the height h of the aircraft.

In practice it is not possible to increase the transmitter frequency continuously in one direction only. In a practical system the frequency is made to deviate about a central carrier

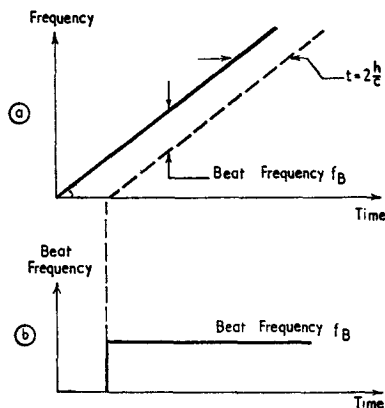


FIG 3. PRINCIPLE OF FMCW RADAR ALTIMETER

frequency by applying a modulating signal to the transmitter. The carrier frequency varies at a *rate* determined by the *frequency* of the modulating signal and by an *amount* determined by the *amplitude* of the modulating signal. The modulating waveform may be triangular, sawtooth, sinusoidal, or some other shape. Most f.m.c.w. radar altimeters use sinusoidal frequency modulation because it is easier to obtain than other methods. Fig 4 shows the result of frequency-modulating a radar altimeter with a sinusoidal modulating signal of frequency f_m . In Fig 4c the solid line indicates the deviation of the transmitter frequency over a band of frequencies B about the centre carrier frequency. The received signal (dashed line) has the same frequency deviation but it is *delayed in time* relative to the transmitted signal by an amount depending upon the height h of the aircraft. Fig 4d shows that the beat frequency f_B produced as a result of applying transmitted and received signals to a mixer is *not constant* over the modulation cycle. At times such as x and y in Fig 4d there is no difference in frequency between transmitted and received signals and the beat frequency is consequently zero at these points. However, if we measure the *average* beat frequency over the period of a modulation cycle this measurement can be used to give an indication of the height h at which the aircraft is flying.

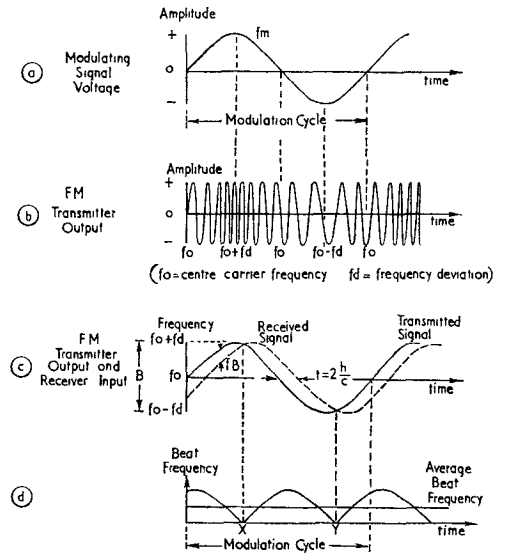


FIG 4. EFFECT OF SINUSOIDAL MODULATION OF FMCW ALTIMETER

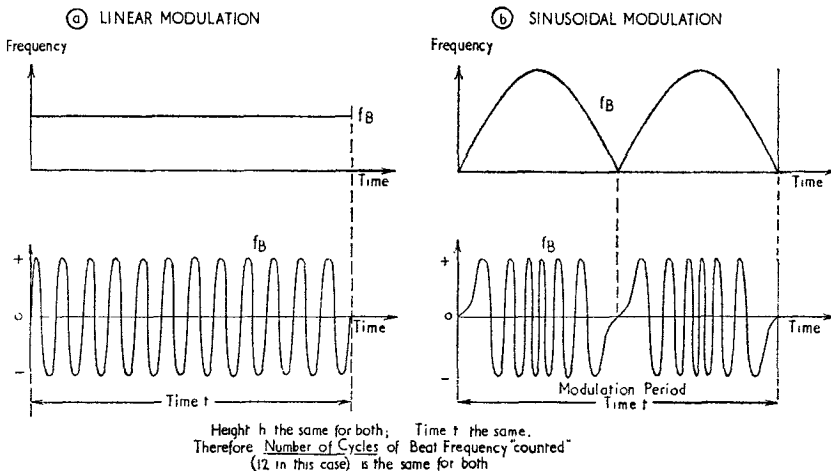


FIG 5. MEANING OF 'AVERAGE' BEAT FREQUENCY

Fig 5 shows what is meant by the 'average' beat frequency. Although the beat frequency f_B varies in a sinusoidally-modulated altimeter as shown in Fig 5b, the *number of cycles* occurring during the modulation period is the same as that which would appear for a linearly-modulated altimeter operating at the same height over the same time period (Fig 5a). By counting the number of cycles in a modulation period we thus *average* the beat frequency over this period to obtain a measure of the height h of the aircraft. A frequency counter is used for this (see p 472).

We have seen that the time taken for a signal to travel from the aircraft to the ground and back is:

$$t = 2 \frac{h}{c}$$

Fig 4 shows that during the period of one cycle of modulation frequency the transmitter frequency undergoes *two* variations of total deviation B . Thus, in one second the transmitter frequency changes by $2Bf_m$ cycles, and in t seconds by $2Bf_m t$ cycles. The average beat frequency f_B is equal to the *change* in transmitter frequency during the *time* taken for the indirect signal to reach the ground and return. Thus, by substituting $2 \frac{h}{c}$ for t we get an average beat frequency of:

$$f_B = \frac{4Bf_m h}{c}, \text{ where } h = \text{height of aircraft above the ground.}$$

B = total deviation of carrier frequency.

f_m = modulation frequency.

c = velocity of radio waves.

Again, the units used must be consistent. If h is in ft, B is in c/s, f_m in c/s, and c in ft/sec. If h is in m, B is in c/s, f_m in c/s, and c in m/sec.

Since the factors B , f_m and c are constant, the average beat frequency f_B is a direct measurement of the height h of the aircraft and is used to produce a direct current proportional to aircraft height.

The height current produced by the altimeter is applied to an altitude indicator. It is also applied to the aircraft automatic pilot where it may be used to fly the aircraft at a predetermined height. In addition, when the aircraft is landing, the height reading is changing rapidly with respect to time (typically 10 ft per second on the glide path). It is possible therefore to compute the *rate of change of height* from the altimeter output and to use this to produce a smooth 'flare' for automatic landing.

Basic Block Diagram of FMCW Radar Altimeter

A block diagram illustrating the principle of operation of the f.m.c.w. altimeter is shown in Fig 6. The arrangement illustrated is typical of the type of altimeter used in the RAF. It operates in the band 4,200 - 4,400 Mc/s and has a nominal unmodulated carrier frequency of 4,300 Mc/s. (We may come across other f.m.c.w. altimeters which work in the frequency band 1,600 - 1,660 Mc/s). The altimeter shown is used to measure heights up to 5,000 ft in two ranges—0 - 500 ft and 0 - 5,000 ft. The frequency deviation about the centre carrier frequency of 4,300 Mc/s depends upon the height range selected; it is ± 5 Mc/s for the high range, and ± 50 Mc/s for the low range. The carrier deviation B is increased for the low range (low value of h) to keep f_B high (see earlier). This gives increased accuracy on the low range. The output is a current proportional to aircraft height and this is applied to the altitude indicator and to the automatic pilot.

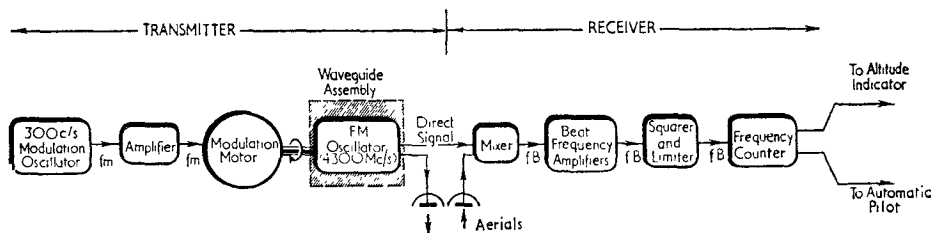


FIG 6. BASIC FMCW RADAR ALTIMETER BLOCK DIAGRAM

Transmitter

The transmitter portion of the altimeter consists of a 300 c/s modulation oscillator followed by an amplifier, the output of which drives a two-phase modulation motor. The motor shaft drives a variable capacitor, which is located in the resonant cavity of the f.m. oscillator, and the variation of the resonator capacitance provides the frequency modulation of the transmitter output.

FM oscillator. This is the 'heart' of the transmitter and provides the f.m.c.w. output at the nominal frequency of 4,300 Mc/s. To supply the necessary transmitter power, which need only be of the order of one watt, different types of microwave oscillators may be used, including the klystron, c.w. magnetron, and backward-wave oscillator (see Sect 4). Another possibility is the velocity-modulated coaxial line oscillator (Fig 7). A coaxial line is short-circuited at one end, the outer conductor being opened out into a disc at the other end, whilst the inner conductor ends in a probe. The inner conductor is split at point A and broadened along a certain portion of its length. The outer conductor has two slits B cut in line with the hollow part of the inner conductor. The electron gun, consisting of cathode, grid and screen, emits a beam of electrons which is focused by a permanent magnet system to form a narrow beam. This beam passes through the buncher and catcher gaps formed by the portions A and B of the coaxial line and is collected by the anode. The resonant line is thus shock-excited into oscillation by the electron beam passing through the gaps and the line resonates at a frequency determined by its length. The coaxial probe transfers energy from the line to a waveguide resonant cavity so that oscillations at the correct frequency (4,300 Mc/s) are maintained. Two outputs are taken from the resonant cavity: a fixed loop is used to couple energy to the transmitter aerial via a small length of coaxial cable; and an adjustable probe extracts a small amount of transmitter energy as the direct signal to the receiver.

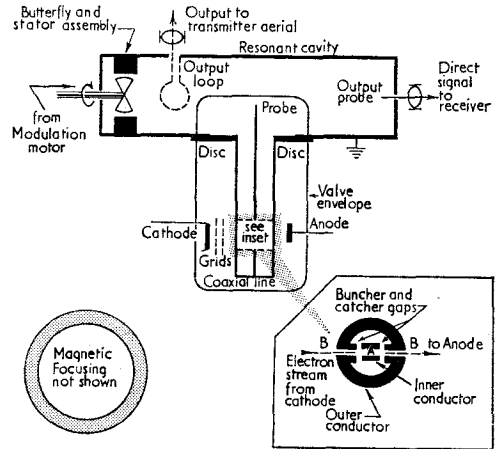


FIG 7. VELOCITY-MODULATED COAXIAL LINE OSCILLATOR

Modulator. The modulation oscillator is a conventional Hartley oscillator, fixed tuned to provide an output at 300 c/s. This output is amplified to provide sufficient drive for the modulation motor. The motor causes a double-vane butterfly assembly to rotate between the vanes of stators mounted in the waveguide resonant cavity (Fig 7). This rotation of the butterfly produces a variation of the waveguide capacitance and causes the transmitter frequency to deviate either side of its nominal carrier frequency. The rate of deviation is 300 c/s and the amount of deviation depends upon the extent of the variation of the waveguide capacitive component. On the high height range (0 - 5,000 ft) the position of the butterfly and stator assembly in the waveguide is such that the frequency deviation caused by rotation of the butterfly is ± 5 Mc/s about 4,300 Mc/s. On the low range (0 - 500 ft) a greater deviation is required (± 50 Mc/s) to provide higher accuracy. To achieve this the butterfly and stator assembly is moved bodily further along the axis of the waveguide resonant cavity to a point of higher electric field strength. The resultant capacitive variation in the guide is then greater. The movement is done automatically by means of actuators controlled by the range switch on the altimeter control unit.

Other methods of modulation are possible. In altimeters using klystrons the frequency deviation may be obtained electronically by varying the reflector voltage. Magnetrons may be electronically frequency-modulated, or they may be mechanically modulated by vibrating an internal reed assembly which varies the capacitance across the straps of the cavity.

Receiver

The receiver has two inputs: the indirect signal from the ground via the receiver aerial; and the direct signal from the transmitter resonant cavity. The function of the receiver is to combine the two inputs to produce a beat frequency equal to their frequency difference; to amplify this beat frequency; and to 'count' the number of cycles in the beat frequency so as to produce an output direct current proportional to aircraft height. The block diagram of Fig 6 shows that the receiver consists basically of a crystal mixer, a low-frequency amplifier, a squarer and limiter, and a frequency counter.

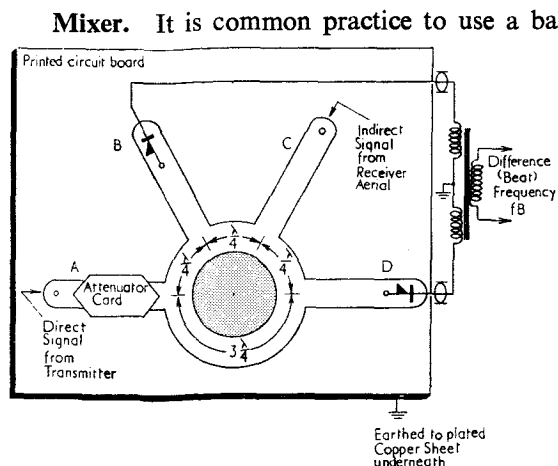


FIG 8. HYBRID RING BALANCED MIXER, STRIPLINE CONSTRUCTION

direct signal; there is no feed from arm A into arm C.

Similarly, energy fed into arm C from the receiver aerial (the indirect signal) divides round the ring to supply *in-phase* inputs to the two crystals in arms B and D; there is no feed from arm C into arm A.

By normal balanced mixer action a beat frequency output is developed across the secondary of the output transformer. This beat frequency is equal to the difference in frequency between direct and indirect signals and is therefore proportional to aircraft height. Since the direct signal drives the crystals in antiphase, much of the transmitter noise present in the direct signal is cancelled in the primary of the push-pull output transformer.

Low-frequency amplifier. The function of the low-frequency amplifier unit following the mixer is to amplify the beat frequency output from the mixer to a level adequate to operate the squarer and limiter circuits whilst maintaining the required signal-to-noise ratio. It will be remembered that the beat frequency output from the mixer has the form shown in Fig 5b. Although the beat frequency varies between zero and a maximum value, it is the number of cycles to be counted by the frequency counter over a modulation period which we consider to be the average beat frequency. This 'averaged' beat frequency will be within the range 100 c/s to 150 kc/s, and of amplitude from about one volt down to 200 microvolts. The frequency and amplitude of the beat frequency signal depend upon the *height* at which the aircraft is flying. At low altitudes we can expect a strong signal at low frequency, whilst at high altitudes the signal will be weaker and much higher in frequency (Fig 9). It is therefore usual to adjust the frequency response characteristic of the amplifier to provide attenuation at the low frequencies corresponding to low altitudes and strong input signals. Less attenuation is applied at the higher frequencies where the input signals are weaker because of the higher altitude.

Mixer. It is common practice to use a balanced mixer as the first stage of the receiver (see p306). This reduces the transmitter noise present in the direct signal. A typical arrangement uses a balanced hybrid ring mixer (see p300). The hybrid ring may be formed by using stripline instead of waveguide or coaxial cable. In this a photo-etched copper microstrip is deposited on a PTFE dielectric sheet (Fig 8). The back of the sheet is plated with copper over its whole surface to provide an effective ground plane.

The direct signal from the transmitter resonant cavity is applied to arm A which contains an attenuator card of graphite-loaded fibreglass to reduce the direct signal to the required level. Energy fed into arm A divides equally in the ring and, by considering the path lengths, it may be seen that the two crystals in arms B and D are driven *in antiphase* with each other by the

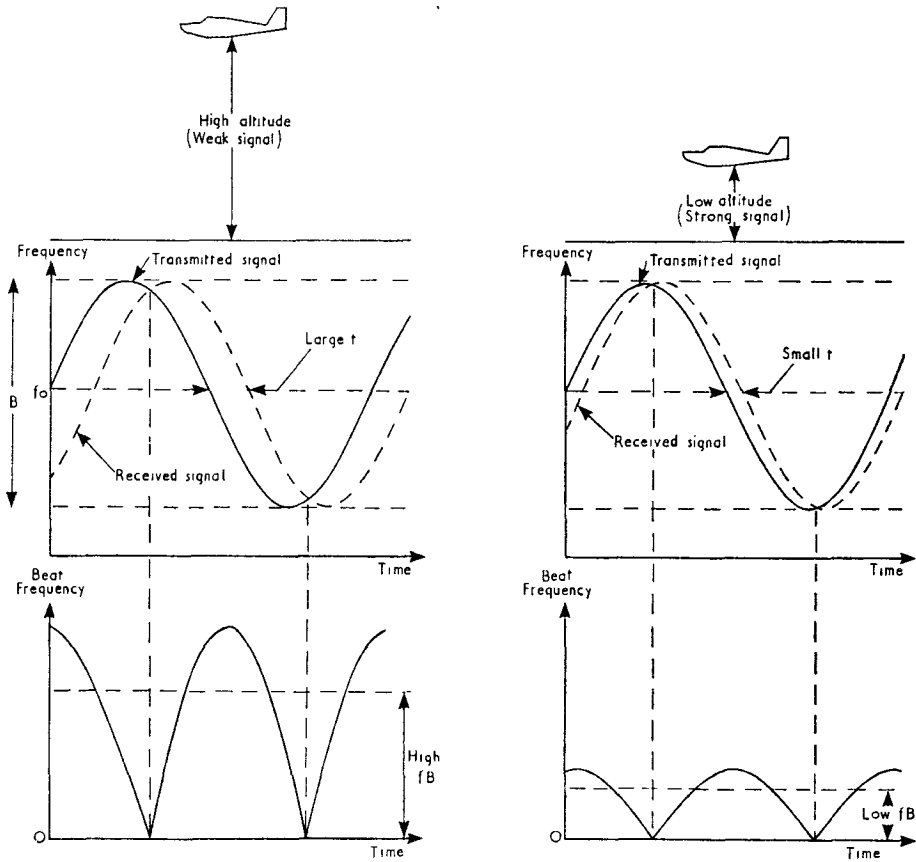


FIG 9. EFFECT OF ALTITUDE ON BEAT FREQUENCY AND STRENGTH OF SIGNAL

Squarer and limiter. For the frequency-counter circuit to operate satisfactorily the input to it must be a good square wave with steep leading and trailing edges and unvarying in amplitude. This waveform may be obtained by applying the sinusoidal beat frequency output of the low-frequency amplifier unit to an Eccles-Jordan bistable circuit (see p127) or to a Schmitt trigger (p131), and then applying the resulting square wave output to a limiter to remove any amplitude variations. The output is then a constant-amplitude square wave whose averaged frequency varies with aircraft height.

Frequency counter. The basic principles of frequency counting circuits are dealt with in p170. The purpose of a counter in a radar altimeter is to provide a current analogue of aircraft height; that is, it produces an output current which is proportional to the averaged frequency of the square wave beat signal applied to the counter.

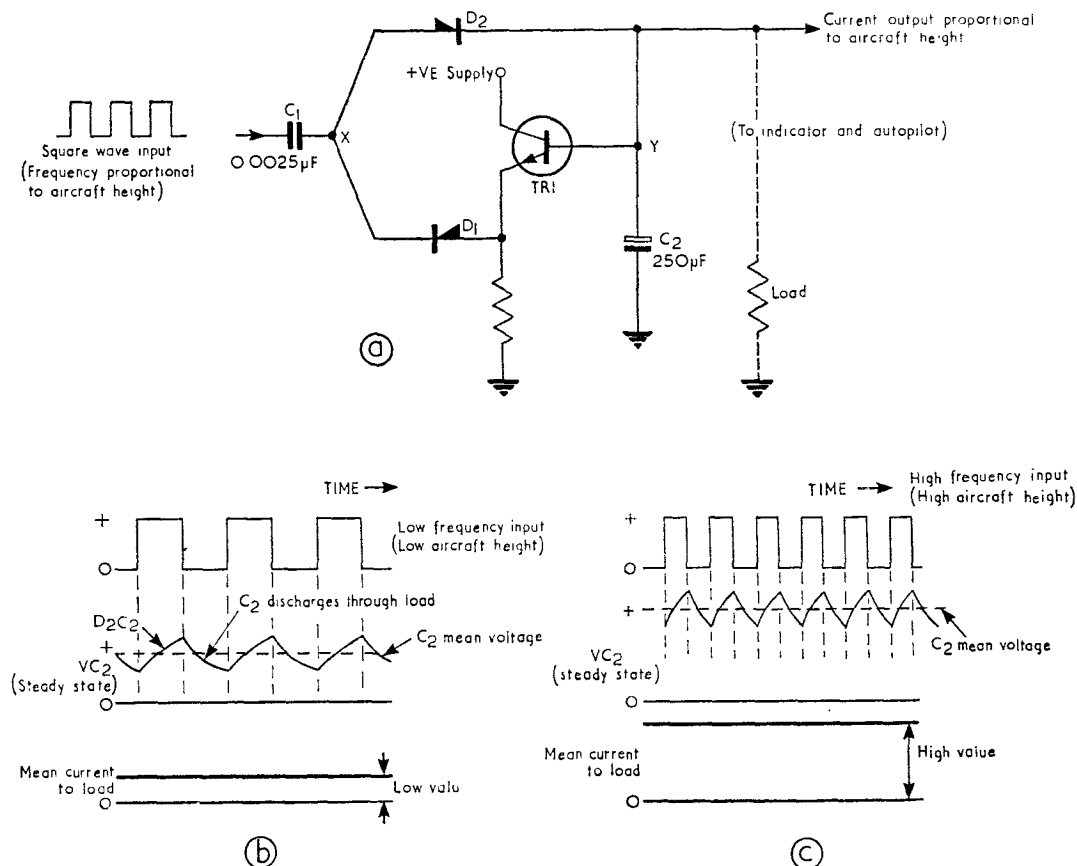


FIG 10. TYPICAL FREQUENCY COUNTER, CIRCUIT AND WAVEFORMS

The circuit and waveforms of a typical counter are illustrated in Fig 10. The input to the counter has a square waveform where the number of cycles in the modulation period is proportional to aircraft height. The leading edge of the first input pulse is applied through C_1 to D_2 causing the diode to conduct. Both C_1 and C_2 then charge rapidly through D_2 until $V_{C1} + V_{C2}$ equals the applied voltage. At the end of the pulse the voltages across C_1 and C_2 are such that D_2 cuts off whilst D_1 cuts on to discharge C_1 ; C_2 in the meantime starts to discharge slowly through the load (indicator and autopilot).

The leading edge of the next input pulse again cuts on D_2 and C_2 acquires a further charge, C_1 also recharging. At the end of the pulse D_2 cuts off and D_1 again conducts to discharge C_1 ; C_2 again discharges slowly through the load.

Continuing this process we see that C_2 acquires a charge during the pulse, some of this charge leaking away through the load during the inter-pulse periods. After a time the charge gained by C_2 via D_2 is exactly counter-balanced by that lost by C_2 via the load. The *mean voltage* across C_2 is then constant for that particular input (Fig 10b).

If the number of cycles of the square wave input during the modulation period *increases* due to an increase in aircraft height, C_2 has less time in which to discharge during the inter-pulse periods and C_2 mean voltage reaches a *higher* level (Fig 10c).

The current output to the indicator and automatic pilot depends upon the mean voltage of C_2 and is thus proportional to aircraft height.

The function of the transistor TR_1 is similar to that described in p172 for the transistor counter. Without TR_1 the charge on C_2 would bias the diode D_2 and the input would then have to exceed this bias before D_2 could conduct to re-charge C_2 . TR_1 prevents this by effectively clamping point X to point Y. Since X is at nearly the same voltage as Y, D_2 has practically no reverse bias to overcome and the whole input voltage is thus available during each pulse to maintain the charge on C_2 .

Limit Lights

Most f.m.c.w. radar altimeters have a 'limit light' system. This consists of three lights, normally coloured amber, green, and red, mounted on the pilot's instrument panel. They are used to assist the pilot to fly the aircraft at a pre-determined and pre-selected height. If the pilot wishes to fly at a height of 400 ft he selects this height by means of a switch on the altimeter control unit. He then adjusts the aircraft flight controls until the aircraft is flying at a height of 400 ft as shown on the altitude indicator. When the aircraft is at this height, the green (ON) light switches on. If the height of the aircraft above the ground deviates from this point one of the other lights comes on. If flying *below* the pre-selected height, the red (LOW) light comes on; if *above* the required height, the amber (HIGH) light comes on.

One arrangement for providing the necessary switching is illustrated in Fig 11. The output from the counter is a direct current which is proportional to the height of the aircraft. This current is applied to the base of an emitter-follower, together with a reference current whose value depends upon the height pre-selected at the control unit by the pilot. The emitter-follower operates two relays $\frac{RLA}{I}$ and $\frac{RLB}{I}$. If the aircraft is flying higher than its pre-selected

height the output from the emitter-follower is sufficient to energize both relays; contacts $RLA1$ and $RLB1$ both close to the ON position and the amber (high) light comes on. If the height of the aircraft is reduced so that it is now flying at the correct pre-selected height, the counter and reference inputs balance and the output from the emitter-follower falls sufficiently to de-energize $\frac{RLB}{I}$ whilst $\frac{RLA}{I}$ remains on. With contacts $RLB1$ OFF and $RLA1$ ON,

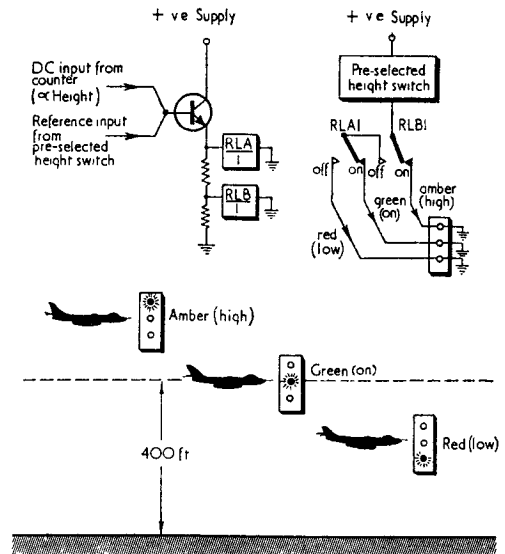


FIG 11. LIMIT LIGHT OPERATION

the circuit to the green (ON) light is complete. If the aircraft height falls below that pre-selected, the emitter-follower is biased off by the reference input (which now exceeds the input from the counter) and both relays are de-energized. With contacts RLA1 and RLB1 both OFF, the red (low) light is switched on.

Aerials

In older types of radar altimeters the transmitting and receiving aerials were dipoles with reflectors. In modern altimeters, to produce the required beam-width, the aerials are wave-guide horns fitted flush with the aircraft skin. The required beamwidth is critical. If the beamwidth is too wide the reflected signal strength is reduced and there is also the tendency for the altimeter to be affected by obstructions near the runway during the flare phase of a landing. On the other hand, if the beamwidth is too narrow, the height measurement will be affected by any variation in the pitch or roll of the aircraft. Fig 12 shows that with a very narrow beamwidth the altimeter reads *slant range* to the ground and so gives an incorrect indication of height. In practice a beamwidth of about 40° to the half-power points is used. With this beamwidth the altimeter gives the correct height reading for angles of $\pm 20^\circ$ in the pitch and roll attitudes of the aircraft.

Typical Equipment

Fig 13 illustrates in diagrammatic form a typical f.m.c.w. altimeter as used in the RAF. It indicates the height of an aircraft up to a range of 5,000 ft above any type of terrain. The height measurement is covered in two ranges—0-500 ft and 0-5,000 ft.

The lower range is intended to give height guidance in the final stages of an automatic landing. The altimeter operates within the frequency band 4,200 - 4,400 Mc/s.

The measurement accuracy of a radar altimeter depends to a large extent on the total frequency deviation used. A *large* frequency deviation reduces the errors in altimeter reading. Thus, on the low range, where accuracy is most important, the frequency deviation is ± 50 Mc/s. On the high range the deviation is reduced to ± 5 Mc/s. On the low range the error does not exceed ± 3 ft or $\pm 3\%$ of height, whichever is greater; at touchdown the error does not exceed 1 ft. On the high range the error does not exceed ± 30 ft or $\pm 3\%$ of height, whichever is greater. The aerial beamwidth is such that variations in pitch or roll up to $\pm 20^\circ$ are allowed for. The power output into the aerial is 0.5 watt.

The Altimeter in Automatic Landing Systems

To maintain aircraft movements under all types of weather conditions it is becoming increasingly important to provide aircraft with a completely automatic landing facility. In any system of this type, the radar altimeter has a large part to play. Automatic landing systems are considered briefly in p243 of Part 2 of these notes. The instrument landing system (ILS) forms an integral part of any automatic landing system and is used to provide both horizontal and vertical guidance during the approach phase to a runway. During this period the radar altimeter reading gives a check on height. On nearing touchdown, however, an accurate and

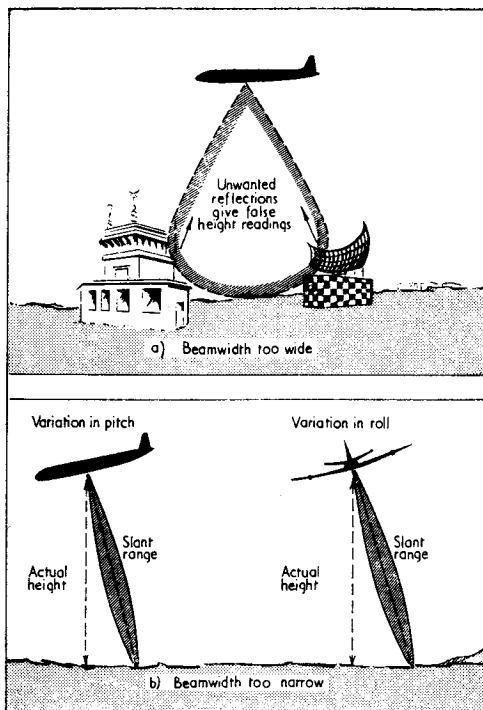


FIG 12. EFFECT OF AERIAL BEAMWIDTH

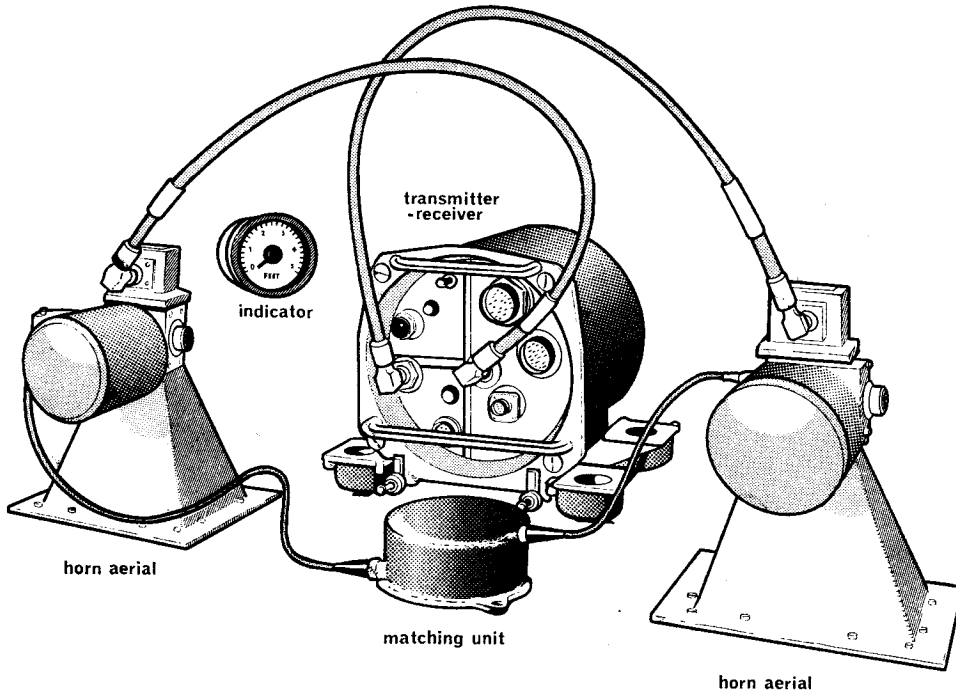


FIG 13. LAYOUT OF TYPICAL FMCW RADAR ALTIMETER

continuous measurement of aircraft height above the runway is necessary to control the final flareout. For this purpose the radar altimeter is used specifically to provide height information to the automatic pilot from about 60 ft down to touchdown.

A typical automatic landing is illustrated in outline in Fig 14. The initial part of the landing, down to about 150 ft, is controlled by the ILS glide slope path. From 150 ft down to 60 ft, the memory in the computer of the automatic pilot 'remembers' this information and maintains the aircraft glide path. The actual height is continuously monitored by the altimeter. From 60 ft down to touchdown the radio altimeter output is switched in to control the automatic pilot. The height measurement is computed to provide a *rate of change of height* factor which, in turn, determines the flare-out path.

In such a critical operation it is necessary to ensure that the height measurement is continuously available. In modern aircraft this is assured by duplicating—or in some cases, triplicating—the radar altimeter equipment. In the event of failure of one equipment, an immediate change-over to a standby equipment takes place.

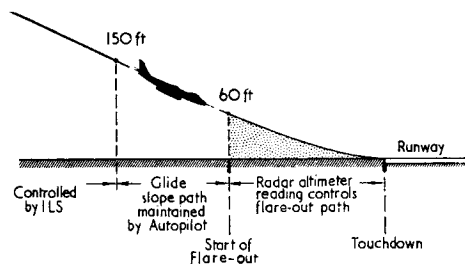


FIG 14. RADAR ALTIMETER IN AUTOMATIC LANDING

Modern Developments in Radar Altimeters

From a radar altimeter we require high accuracy and good reliability. The first is obtained by good design of the equipment and, as we have seen, modern altimeters have an accuracy of about 1 ft at touchdown. Reliability can be improved by substituting solid state devices for thermionic devices where practicable. In some present-day altimeters every stage uses tran-

sistors or semiconductor devices with the exception of the transmitting valve. This gives good reliability. However, the continued search for improved reliability will obviously take advantage of the thin-film and integrated solid circuits being developed. Thin-film and integrated circuits are considered later in this book. It is sufficient at this stage to say that a thin-film circuit consists of photo-etched layers of conducting and dielectric materials placed on a substrate to form the 'passive' parts of a circuit (e.g. resistors and capacitors); micro-miniature transistors may then be connected to the thin-film strips as the 'active' elements. In integrated circuits various layers are diffused into a solid 'chip' of semiconductor material to fabricate a completely integrated circuit within the chip. The reduction in the numbers of interconnections both in thin-film and integrated circuits gives improved reliability. Thus, future radar altimeters will be all-solid-state employing either thin-film or integrated circuits or a combination of both. The use of such circuits will provide a valuable bonus: a remarkable reduction in both size and weight of equipment.

AIRBORNE DOPPLER PRINCIPLES

Introduction

With large numbers of high-speed aircraft operating in crowded air-spaces, it is important that each aircraft be fitted with a self-contained navigation system that is capable of operating under any weather condition. Such a navigation system should provide the necessary information for accurately piloting the aircraft from one position to another without the assistance of a ground station. Airborne Doppler radar provides this information.

In aircraft navigation, two basic questions have to be answered: the first is concerned with the time of arrival of the aircraft at its destination; and the second concerns the heading to be steered to reach the required destination. Doppler radar measures the speed of the aircraft over the ground beneath (the *groundspeed*) and so provides information to answer the first question.

The answer to the second is complicated by the fact that it is not sufficient to draw a straight line between two points A and B and set the aircraft on this heading. If no account has been taken of the effect of the wind the aircraft will certainly not finish at B. Just as a rowing boat traveling in a straight line to a point directly across a river has to be headed slightly upstream to counteract the effects of river current, so an aircraft flying on a specified course has to be headed slightly into the wind. The direction in which the aircraft is pointing is termed the *aircraft heading*, and the angle between the aircraft heading and the actual aircraft path over the ground (*track*) is known as the *drift angle* (Fig 1).

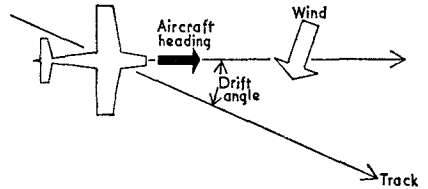


FIG 1. DRIFT ANGLE IN AIRCRAFT FLIGHT

For an aircraft in flight the drift angle is varying continuously in a random manner as the wind speed and direction changes, and as the aircraft speed and heading changes. Thus to keep the aircraft on the required heading, continuous measurement of the drift angle is necessary so that corrections to the aircraft flight path may be made. Airborne Doppler radar may be used to provide continuous measurement of drift angle, as well as groundspeed measurement.

Measurement of Groundspeed

The basic Doppler principles discussed in Chapter 1 of this section for ground c.w. radar apply also to airborne Doppler (see p446). Whether the transmitter-receiver is stationary and the target moving as in ground radar, or the transmitter-receiver moving and the target stationary as in airborne Doppler, the results obtained are basically the same. In either case it is *relative velocity* between source and target with which we are concerned, and the frequency of the received signal differs from that transmitted by an amount equal to the Doppler shift f_D as given in p446:

$$f_D = 2 \frac{v_r f_T}{c},$$

where v_r = component of velocity along sight-line of beam.

f_T = transmitted frequency.

c = velocity of e.m. waves.

In this expression f_D is usually given in c/s (or kc/s), v_r in m.p.h. (or knots), f_T in c/s (or Mc/s), and c in m.p.h. (or knots).

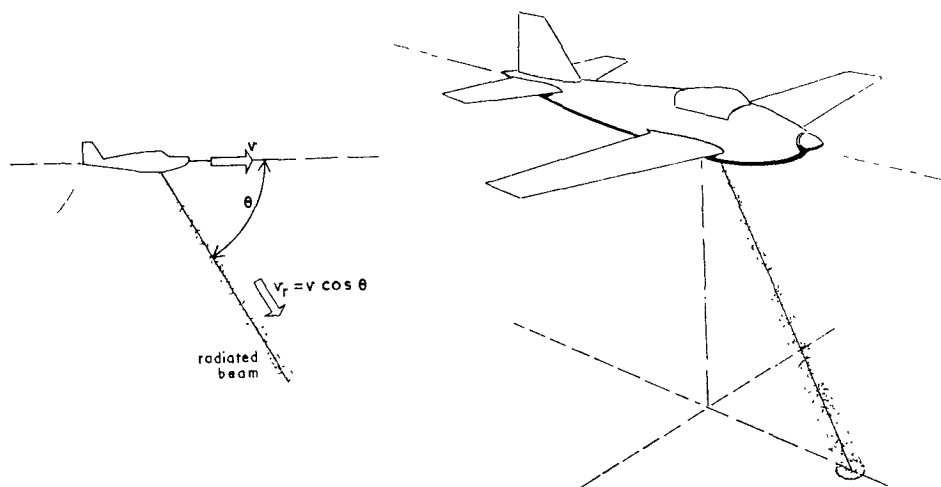


FIG 2. BASIC CONCEPT OF GROUNDSPD MEASUREMENT

Fig 2 illustrates an aircraft moving with horizontal velocity v relative to the ground. The aircraft is radiating a narrow beam of radio energy at an angle θ to its horizontal direction of motion. The component of velocity along the beam is $v_r = v \cos \theta$, and the magnitude of the Doppler shift in the ground return is now:

$$f_D = 2f_T \frac{v \cos \theta}{c}$$

With the exception of the aircraft groundspeed v , all the factors on the right-hand side of this expression are known. Thus, measurement of the Doppler shift f_D gives a direct measure of the aircraft groundspeed v .

Need for High Operating Frequency

We saw in Chapter 1 that the Doppler shift is approximately 30 c/s per 100 Mc/s transmitted frequency per 100 m.p.h. relative velocity, where relative velocity equals $v \cos \theta$. The choice of depression angle θ is a compromise. If θ is too small, the echo received back at the aircraft will be weak. On the other hand, if θ is large, its cosine is small and the value $2f_T \frac{v \cos \theta}{c}$ may become too small for accurate measurement. Most airborne Dopplers have angles of beam depression θ of the order of 60° to 70° ; if $\theta = 60^\circ$, $\cos \theta = 0.5$. It is then necessary to make the transmitter frequency f_T high so that the beat frequency produced by mixing the Doppler-shifted echo with a portion of the transmitted output can be easily measured.

Two frequency bands are used for airborne Dopplers: one centred on 8,800 Mc/s, and the other on 13,300 Mc/s. For an 8,800 Mc/s Doppler, the beat frequency obtained at 600 m.p.h. assuming an angle of depression θ of 60° , is of the order of 7.5 kc/s—an audio frequency that can be measured accurately to give an output in terms of groundspeed v .

Limitations of Single Aerial Beam

In what we have learnt so far we have assumed that the aircraft was in straight and level flight. If the aircraft deviates from level flight, the value of the depression angle θ of the beam alters, as in Fig 3a. As we have seen, this causes the Doppler shift f_D to vary, implying that the groundspeed has changed. The horizontal velocity of the aircraft relative to the ground may not, in fact, have changed. Thus, with a single beam, variation in the *pitch* of the aircraft gives rise to errors in groundspeed measurement.

The Doppler shift from a single beam aerial gives the component of aircraft velocity in the *direction of the beam*. Assuming horizontal flight and that the beam is directed straight ahead, this gives the velocity in the direction of the aircraft *heading* (Fig 3b). However, if drift is present, the aircraft heading differs from the track by the drift angle, and inaccurate measurement of groundspeed along the aircraft's *track* results.

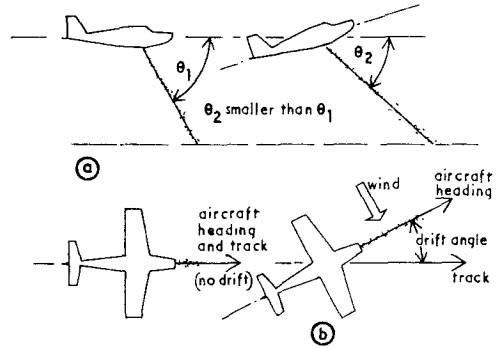


FIG 3. EFFECTS OF VARIATIONS IN AIRCRAFT PITCH AND DRIFT ANGLE ON GROUND SPEED MEASUREMENT

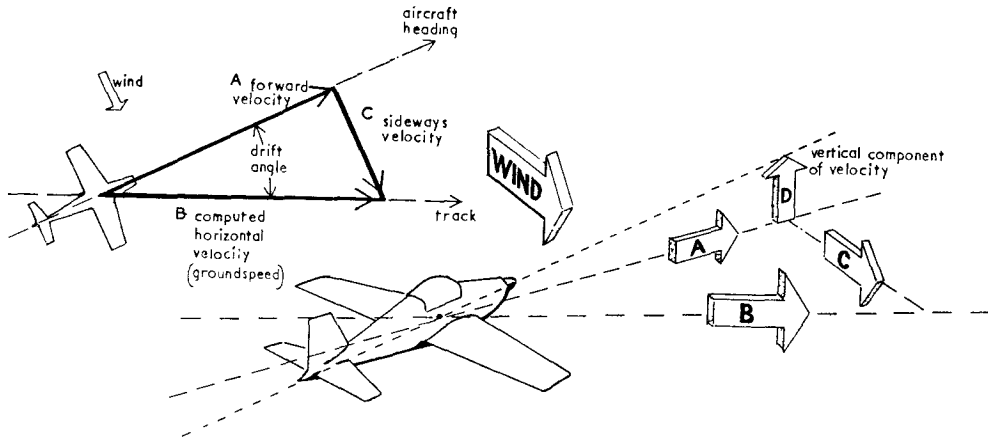


FIG 4. COMPONENTS OF AIRCRAFT VELOCITY

We have, in fact, to consider *three* components of aircraft velocity. If the aircraft is climbing or diving we have a *vertical* component of velocity D (Fig 4). In the horizontal plane we have *two* components of velocity: a *forward* velocity A caused by the aircraft's thrust; and a *sideways* velocity C from the effect of the wind. To determine the speed and direction of travel of the aircraft, all three components of velocity should be computed. In practice, however, the vertical component of velocity is normally dealt with separately and the two components of velocity in the horizontal plane are computed to give B (Fig 4). In this way the drift angle is obtained and the groundspeed along the aircraft's *track* measured. With a single beam system, only the *forward* component of velocity along the aircraft *heading* is measured and, as we have seen, this produces errors in groundspeed measurement.

Use of Forward and Backward Beams

By radiating *two beams*, one directed forward and the other to the rear of the aircraft, errors in groundspeed measurement, caused by vertical motion or variations in the *pitch* of the

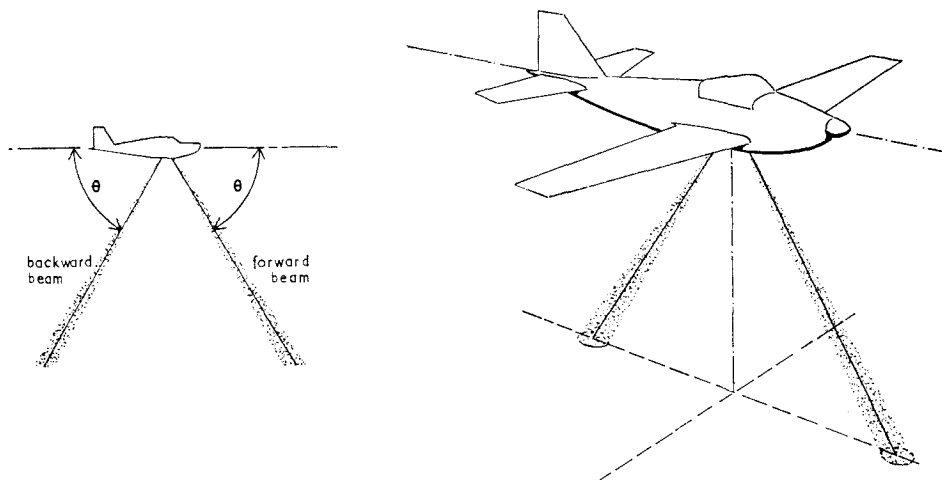


FIG 5. USE OF FORWARD AND BACKWARD BEAMS

aircraft, can be greatly reduced. Fig 5 illustrates the arrangement and shows that both beams are depressed through angle θ with reference to the horizontal axis of flight.

The frequency of the echo signal f_R received by the *forward* beam is *higher* than that of the transmitted signal f_T by an amount equal to the Doppler shift f_D :

$$f_R = f_T + f_D$$

The echo received by the *backward* beam has a *negative* Doppler shift and has a frequency of:

$$f_R = f_T - f_D$$

In some Doppler equipments these two received signals are mixed together and the difference frequency extracted to produce a beat frequency f_B of:

$$f_B = (f_T + f_D) - (f_T - f_D) \\ = 2f_D$$

$$f_B = 4f_T \frac{v \cos \theta}{c} \quad (\text{since } f_D = 2f_T \frac{v \cos \theta}{c} \text{ as given earlier}).$$

This method has the advantage that the beat frequency produced has *twice* the value of that from a single beam so that greater precision in the measurement of aircraft groundspeed is possible. Frequency stability of the transmitter also becomes less important: changes in transmitter frequency affect the echo signals in the two directions equally and are therefore cancelled when taking the difference frequency. However, since the information is extracted by mixing two small echo signals, the system is less efficient than that in which the echo is mixed with a large reference signal (local oscillator or sample of transmitter output).

Let us now suppose that we have a variation in the pitch of the aircraft. The value of the depression angle θ is increased for one beam and decreased *by the same amount* for the other. For *small changes* in pitch the *decrease* in frequency of the Doppler-shifted echo from one beam is, therefore, almost exactly balanced by the *increase* in frequency from the other. Hence the beat frequency produced by mixing the two signals is virtually unaffected by small variations in pitch. For example, a 1° variation in the pitch of an aircraft produces a groundspeed error of about 3% in a single beam system; in the double beam system this error is reduced to 0.02%.

The error in both systems increases as the angle of pitch increases. For variations in pitch greater than about 5° the groundspeed error in the *double beam* system becomes important. For this reason it is normally necessary to stabilize the aerial in the horizontal plane. This may be done by mounting the aerial on a stable platform whose position with respect to the horizontal plane is controlled by a gyroscope (see AP3302 Part 4).

Counteracting Effect of Drift on Groundspeed

We have seen that with a single beam system, velocity is measured along the *aircraft heading* and, in the presence of drift, errors in groundspeed measurement result. This may be overcome by radiating two beams, one to the right and the other to the left of the aircraft, both beams depressed by the same angle θ towards the ground (Fig 6). The aerial array is aligned with the fore-and-aft axis of the aircraft and radiates the two beams *symmetrically* about this axis, each beam making an angle θ with the aircraft centre line.

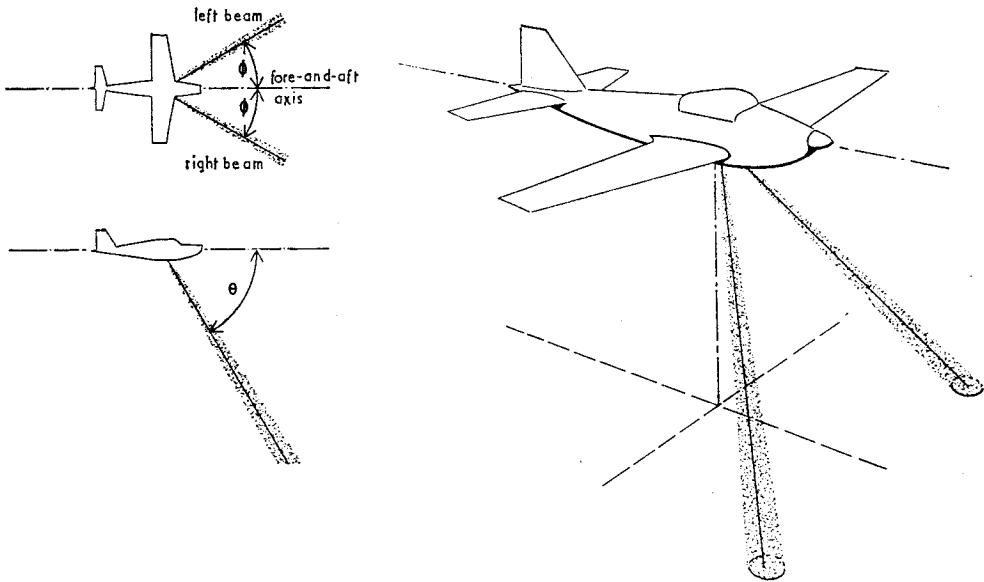


FIG 6. COUNTERACTING EFFECT OF DRIFT ON GROUNDSPED MEASUREMENT

If there is no drift, the aircraft heading coincides with the aircraft track. The symmetrical beams then produce equal Doppler shifts. However, a drift to left or right moves the aircraft heading away from the track and this produces a greater Doppler shift in either left or right beam. A computer could then be used to resolve the difference in frequency into a measure of drift angle. Alternatively, the difference in Doppler shifts can be used as an *error signal* to drive the axis of the aerial array to a position where the error signal falls to zero. This occurs when the two beams are symmetrical about the aircraft *track*. The drift angle is then obtained by direct analogue from the angle the aerial array makes with the fore-and-aft axis of the aircraft. The aerial in this position is aligned with the aircraft track and so accurate measure of groundspeed is possible.

Janus System

Most modern airborne Dopplers combine the forward-backward beam system and the port-starboard beam system to form a *four-beam* system. The arrangement, illustrated in Fig 7, is known as the *Janus* system (after the Roman god who looked both forward and backward at the same time).

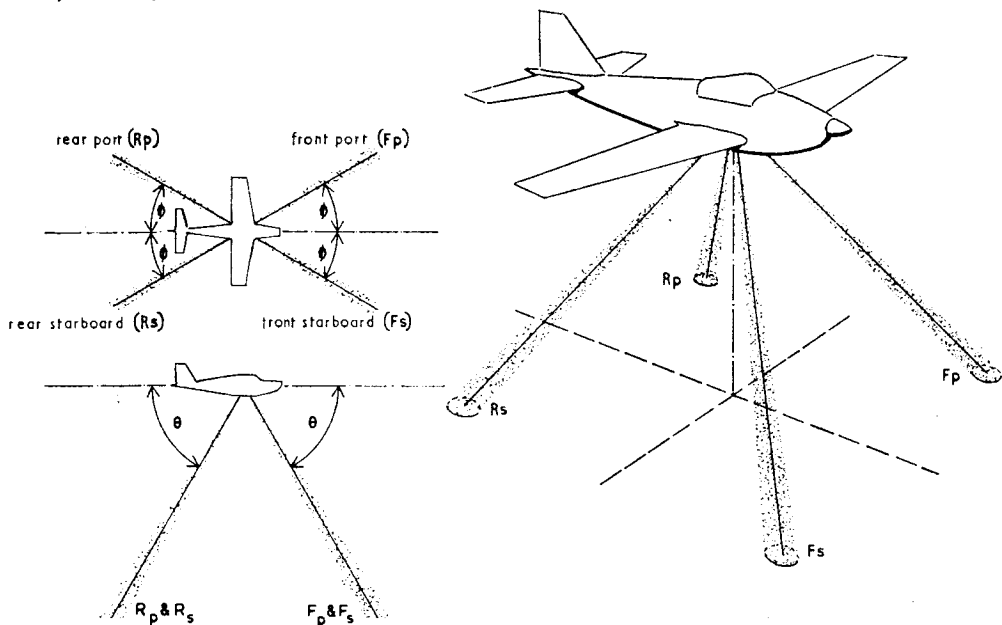


FIG 7. JANUS FOUR-BEAM DOPPLER AERIAL SYSTEM

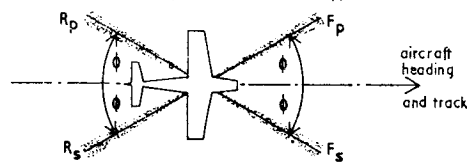
Some airborne Dopplers use a fixed aerial with the four beams directed symmetrically about the fore-and-aft axis of the aircraft. In other installations the aerial is free to rotate in azimuth about the fore-and-aft axis.

a. Fixed aerial. The usual method is to transmit energy simultaneously from beams F_p and R_s . The received echoes are mixed and, as we have seen earlier, the Doppler component is extracted as a beat frequency f_{B1} . This frequency is a measure of the component of aircraft velocity in the direction of the beam F_p . Similarly, a beat frequency f_{B2} is obtained by mixing the echoes from beams F_s and R_p . This frequency is a measure of the component of aircraft velocity in the direction of the beam F_s . Since the angles between the beams and the fore-and-aft axis are known, the difference in beat frequencies can be computed to give a measure of drift angle and ground-speed.

b. Moving aerial. As in the fixed aerial system, the beat frequency f_{B1} from beams F_p and R_s is compared with f_{B2} from beams F_s and R_p . If there is no drift present, the aircraft's track is the same as its heading. Thus, with the beam-pairs disposed symmetrically about the aircraft track, the beat frequencies f_{B1} and f_{B2} will be equal, as in Fig 8a. The groundspeed can then be found directly from the relationship:

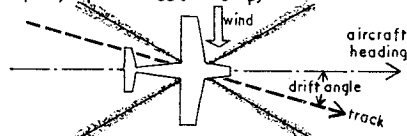
$$f_{B1} = f_{B2} = 4f_T \frac{v \cos \theta \cos \phi}{c}$$

beat frequency $f_{B1}(\text{from } F_p - R_s) = f_{B2}(\text{from } F_s - R_p)$



① No drift—beams symmetrical about track

$f_{B1}(\text{from } F_p - R_s)$ LESS THAN $f_{B2}(\text{from } F_s - R_p)$



② With drift—beams no longer symmetrical about track

FIG 8. MEASUREMENT OF DRIFT AND GROUND-SPEED, JANUS MOVING AERIAL SYSTEM

In Fig 8b the wind has introduced a drift to starboard. The component of aircraft velocity along the beam F_s is now *greater* than that along F_p , so that beat frequency f_{B2} is greater than f_{B1} . This frequency difference is used to generate an *error signal* which, in turn, drives a servomotor, rotating the aerial in azimuth until the frequency difference is reduced to zero. When f_{B1} equals f_{B2} the beams are symmetrically disposed about the *aircraft's track*. The angle through which the aerial has turned is equal to the drift angle and this is displayed on a meter. The groundspeed is given as before by measuring f_{B1} or f_{B2} , and since the aerial is now directed along the aircraft's track, the groundspeed measurement is accurate.

It will be remembered that because of the forward-backward beam arrangement, the Janus system automatically reduces errors in groundspeed measurement caused by variations in the *pitch* of the aircraft. Most Doppler aerals used in the RAF are also stabilized in the pitching plane.

Three-beam System

Some fixed-aerial Doppler installations use a three-beam Janus system. This gives results similar to those obtained from a four-beam system. In the three-beam system we have two forward-directed and one rearward-directed beams as illustrated in Fig 9. The Doppler-shifted echo in each beam is mixed with a sample of the *transmitter* signal to produce *three separate* Doppler shifts. The Doppler signal from F_p is compared with that from R_p to provide the aircraft *forward* velocity. The Doppler shifts from F_p and F_s are compared to provide the aircraft *sideways* velocity. And the Doppler shifts from F_s and R_p are compared to provide the aircraft *vertical* velocity. The three velocity output signals are fed to a computer which provides the required groundspeed and drift measurements and other navigation information.

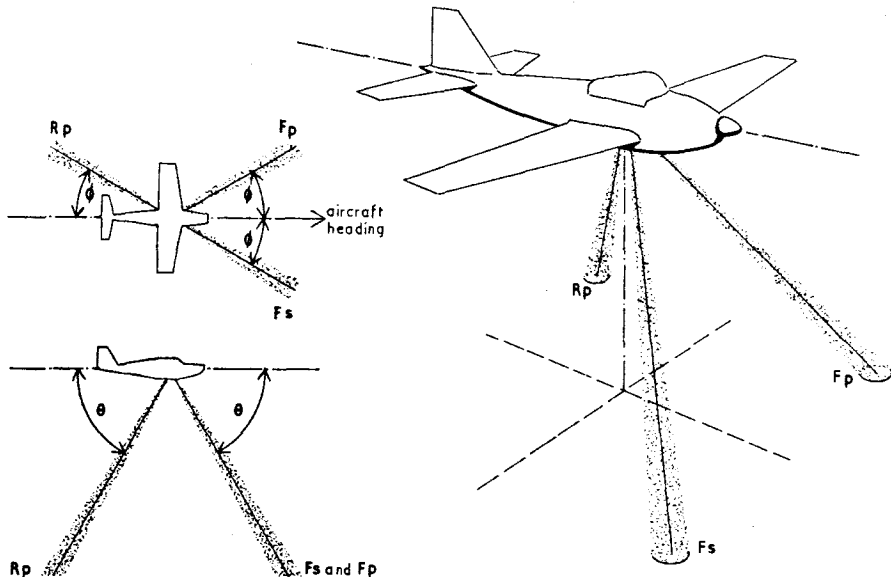


FIG 9. JANUS THREE-BEAM AERIAL SYSTEM

CW Doppler

Three types of transmission are possible in airborne Doppler systems—c.w., f.m.c.w., and pulse. We shall consider each in turn.

In pure c.w. systems the transmitter power is shared equally between each beam. The power output need only be very small and is typically of the order of 100mW. The c.w. system is the simplest of the three types of transmission. It requires no aerial switching, so that losses and unreliability from this cause are reduced; and complex modulation systems, with their

limitations, are avoided (see later). The major difficulty in a pure c.w. system is the *cross-coupling* effect between transmitting and receiving sections of the aerial array, giving rise to errors in measurement. This may be reduced by using *two* aerials—one for the transmitter and the other for the receiver—shielded from each other by the insertion of a metal baffle plate between the two arrays.

The initial stages of a simplified c.w. Doppler arrangement are illustrated in Fig 10. A three-beam, fixed-aerial system is used, with separate aerial arrays for transmitter and receiver.

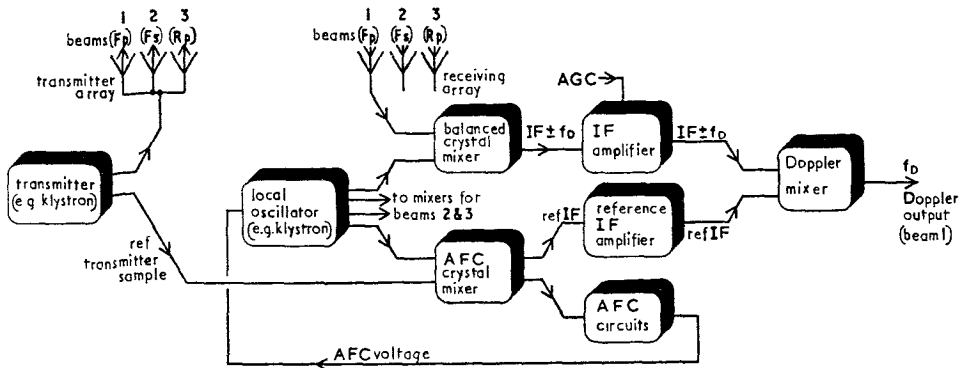


FIG 10. OUTLINE OF AIRBORNE C.W. DOPPLER SYSTEM

The transmitter power is unmodulated and is split equally between the three beams. Since only a low-power output need be provided the transmitter may consist of a low-power klystron, or a crystal-controlled transistor oscillator followed by varactor multipliers. In the receiver we have three channels, all similar, one for each beam. Only one channel is illustrated. The Doppler-shifted echo in each beam is applied to a balanced crystal mixer, each mixer being fed with local oscillator power. Another klystron, or another transistor-varactor chain, may be used for this. The output from the mixer is an i.f. signal on which the Doppler component is superimposed. This signal at $f_{IF} \pm f_D$ is amplified and applied to the Doppler mixer along with the amplified reference i.f. signal from the a.f.c. mixer. The output from the Doppler mixer is the required Doppler shift (see p450). By comparing the Doppler frequencies in the three channels we obtain the forward, sideways and vertical components of velocity as explained earlier.

FMCW Doppler

We saw on p459 that f.m.c.w. radar provides *range discrimination*. Advantage can be taken of this fact to enable a Doppler receiver to be made insensitive to cross-coupling from the transmitter and to echoes from the radome and airframe. At the same time, only a small amount of signal loss from distant returns is introduced.

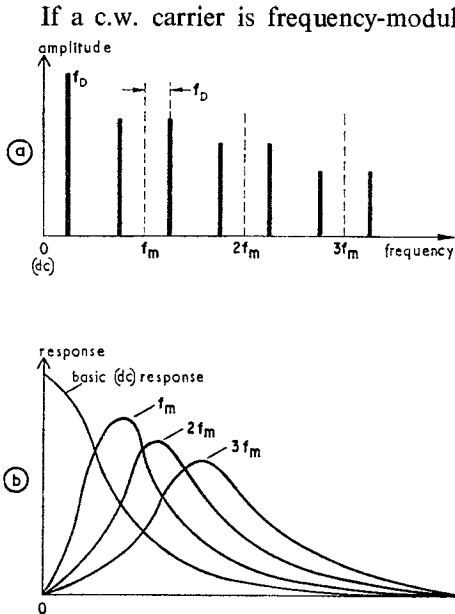


FIG 11. USE OF HARMONICS OF MODULATION FREQUENCY TO REDUCE CROSS-COUPLING EFFECT

returns (at greater range) can be kept high. Those Doppler radars that use f.m.c.w. transmission usually extract the *third* harmonic modulation-frequency component, and the required Doppler signal is obtained from this.

The initial stages of an f.m.c.w. Doppler radar are illustrated in Fig 12. A four-beam, moving-aerial system is used. The f.m.c.w. output of the transmitter is of the order of 1W at a frequency of 8,800 Mc/s (or 13,300 Mc/s). This output may be provided by a klystron, frequency-modulated by a signal of frequency f_m (typically 400 kc/s) produced by a solid state oscillator-modulator. The transmitter output is applied to the aerial array which switches the power to

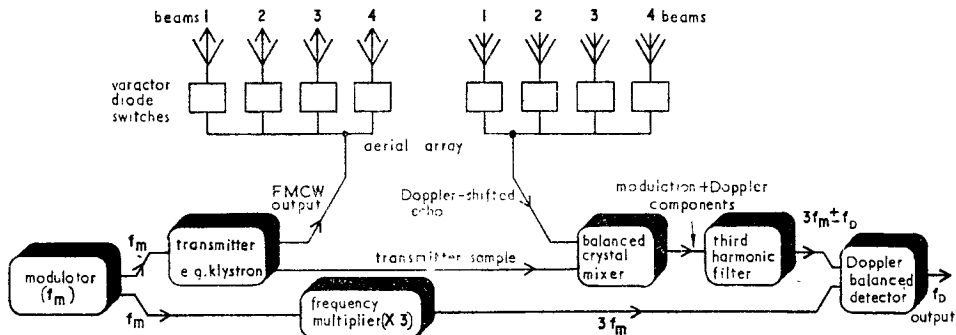


FIG 12. OUTLINE OF AIRBORNE FMCW DOPPLER SYSTEM

each of the four beams in turn on a time-sharing basis to ensure that the maximum energy is radiated by each beam. A portion of the transmitter output is also applied to the balanced crystal mixer in the receiver. The other input to the crystal mixer is the Doppler-shifted echo in each beam. It will be remembered that the Doppler shift in each beam is the *same* when the aerial has been turned through the drift angle into alignment with the aircraft's *track*. The

output from the mixer is then the basic (d.c.) Doppler beat frequency plus the 'sidebands' occurring at harmonics of the modulation frequency f_m . The third harmonic component at $3f_m \pm f_D$ is selected by a filter and amplified before being applied as one input to a balanced Doppler detector. The other input to this detector is the third harmonic of f_m from the frequency-multiplier. Thus the output from the Doppler detector is the required Doppler beat frequency f_D which is selected by a filter and amplified. The Doppler signal is then processed to provide the groundspeed and drift measurements.

Pulsed Doppler

The first successful Doppler radars used pulsed transmission, and this is the system used in most present-day RAF equipment. With pulsed transmission, the problem of cross-coupling between transmitter and receiver is completely overcome because, of course, the receiver is switched off whilst the transmitter is firing. The same aerial may now be used for both transmission and reception, a TR switch being used to connect the aerial to the transmitter for the duration of each pulse, and to the receiver for the interval between pulses. In modern equipments the TR switch may be a ferrite circulator (see p302).

Most airborne pulsed Dopplers use a four-beam, moving-aerial system. Beam-pairs are formed as previously explained: the forward port/rear starboard pair are operative for a given time, the energy then being switched to the other pair. The aerial beam-pattern switching may be carried out by a ferrite gyrator (see p302). Typically, the ferrite switch operates at about 1 c/s so that alternate beam-pairs are operative each half-second. The Doppler shifts in each beam-pair are equal when the aerial is aligned with the aircraft's track. The angle between the aerial axis and the aircraft heading is then equal to the *drift angle*. In this position the Doppler output is a measure of groundspeed along the aircraft's track.

In a typical system the transmitter operates at a frequency of 8,800 Mc/s and produces r.f. pulses of $4 \mu\text{s}$ duration at a p.r.f. of 60 kc/s. The high p.r.f. is necessary for the following reason.

We saw on p404 that in pulsed transmission, the radiated signal consists of the carrier frequency signal plus a large number of 'sidebands'—the usual pulsed radar frequency spectrum. The separation between the sidebands is equal to the p.r.f. as shown in Fig 13a. The frequency of the Doppler-shifted echo differs from that of the transmitter carrier by the Doppler shift f_D , and this is within the range 2kc/s to 25kc/s depending upon the aircraft's groundspeed (Fig 13b). Thus, to keep the Doppler-shifted echo clear of interference from the 'sideband' components of the transmitted pulse, the p.r.f. must be *higher* than the highest expected Doppler frequency. A p.r.f. of 60 kc/s is typical.

It will also be noted that the pulse duration used ($4 \mu\text{s}$) is relatively long. There are two reasons for this. The energy contained in a pulse depends, among other things, upon the pulse duration. By using a long pulse the strength of the received Doppler-shifted echo is increased accordingly. In addition, the bandwidth of a receiver is inversely proportional to the pulse duration, so that by using a long pulse, the required bandwidth is reduced and the noise factor of the receiver improved.

However, by using a long-duration pulse at a high p.r.f., the pulse *duty factor* becomes large (see p28). For a $4 \mu\text{s}$ pulse at a p.r.f. of 60 kc/s the pulse duty factor is approximately 1 : 4. It is only recently that magnetrons capable of operating with this high duty factor, without exceeding the rating of the magnetron, have become available. Typical power outputs are 20W peak and 5W mean power. In earlier pulsed Dopplers, to avoid exceeding the available mag-

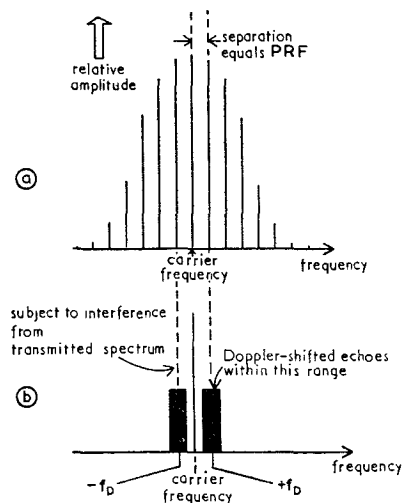


FIG 13. NEED FOR HIGH PRF IN PULSED DOPPLER

netron rating, short-duration pulses had to be used ($0.5 \mu\text{s}$) and even then it was necessary for the magnetron to operate in 'bursts' with adequate resting periods.

The initial stages of a simplified pulsed Doppler arrangement are illustrated in Fig 14. This follows the conventional centimetric pulsed radar pattern and requires no further explanation. The output from the detector is the required Doppler beat frequency within the range 2 kc/s to 25 kc/s, depending upon the aircraft's groundspeed.

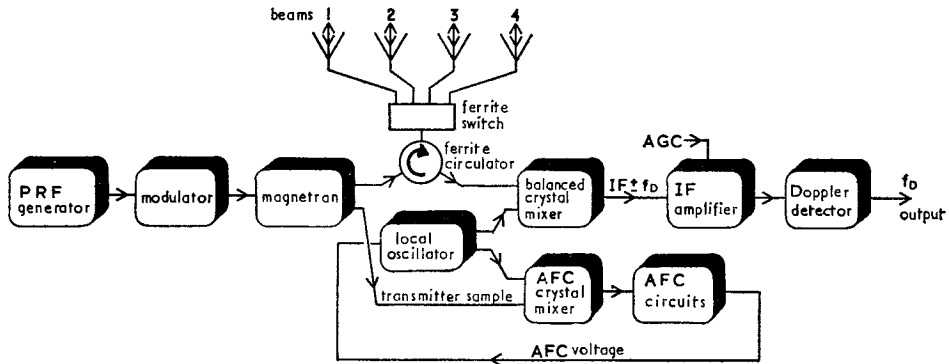


FIG 14. OUTLINE OF AIRBORNE PULSED DOPPLER SYSTEM

Height Holes

Both the f.m.c.w. Doppler and the pulsed Doppler suffer from the effect of height or altitude 'holes'. This effect occurs as a result of the modulation. For a pulsed transmission, if the transmitter fires just when the ground echo arrives back at the radar, the echo will not be detected. Height holes therefore exist at those altitudes (ranges) where the time taken for the signal to reach the ground and return is equal to the time interval between pulses (Fig 15b) or to a *multiple* of that time.

Similar remarks apply to f.m.c.w. Doppler. Here the effect occurs when the aircraft is at such a height that the echo, at the instant of its reception, is in modulation phase with the transmitted signal. The echo then gives the same response as that from a signal at zero range and so is rejected.

We have seen that the p.r.f. in a pulsed system and the modulation frequency in an f.m.c.w. system are both required to be *high*. Thus the time interval T in Fig 15, and the time period of a modulation cycle in f.m.c.w. Doppler, are short enough to give *several* height holes at various aircraft altitudes. At such heights the Doppler information is lost.

Height-hole effect can be reduced by 'jittering' the p.r.f. in a pulsed system, or 'wobbling' the modulation frequency in an f.m.c.w. system. In a pulsed system the p.r.f. may be jittered between the limits of 50 kc/s and 70 kc/s at a rate of 60 c/s. In an f.m.c.w. system the modulation frequency may be wobulated

between 340 kc/s and 460 kc/s at a rate of 10 times per minute. In either case there is now no height at which the Doppler information is completely lost.

Further reduction of height-hole effect can be obtained by using a *fan-shaped beam*. Returns from the ground will then be spread over a considerable range of delays and a portion of the signal will be received at all times—even at critical height-hole altitudes.

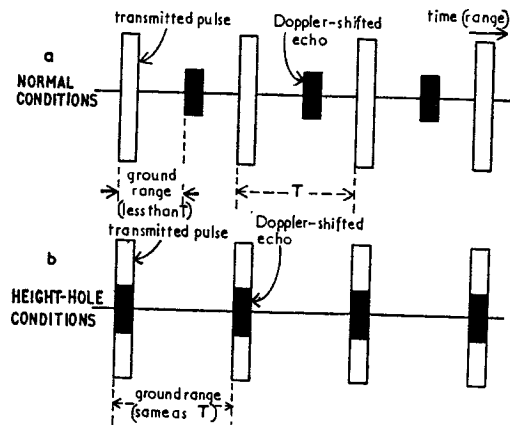


FIG 15. HEIGHT-HOLE EFFECT, PULSED DOPPLER

Beam Shape and Doppler Spectrum

We noted on p 479 that the change in frequency of the Doppler-shifted echo in a beam depends upon the value of depression angle θ of the beam ($f_D = 2f_T \frac{v \cos \theta}{c}$). However, this

assumes that the beam has no 'width' and that the echo is reflected from a *single point* on the ground. In practice, the transmitted beam must have a finite width and, therefore, the echo is made up of the *vector sum* of signals from a large number of reflecting points. Most airborne Doppler radars operate with a depression *beam-width* of about 4° , measured at the half-power points. Thus, for a depression angle of 60° and a depression beamwidth of 4° , the angle θ is 'spread' over the range 58° to 62° (Fig 16a). Instead of a single Doppler frequency shift the echo now contains a *spectrum* of Doppler frequencies ranging from $2f_T \frac{v \cos 62^\circ}{c}$ to $2f_T \frac{v \cos 58^\circ}{c}$. This is indicated in the graph of Fig

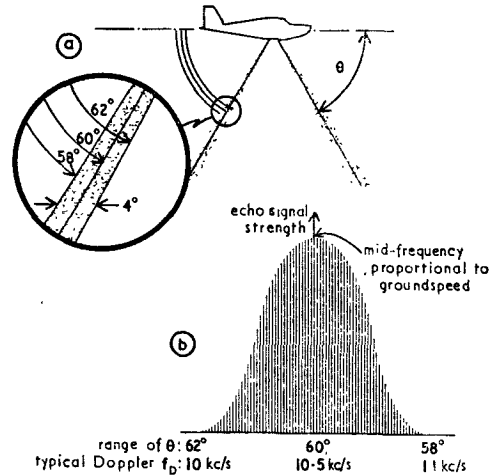


FIG 16. DOPPLER SPECTRUM

16b. In airborne Dopplers a *frequency tracker* is used to measure the mid-point of this spectrum to obtain groundspeed. We shall see later how this is done.

Sea Bias

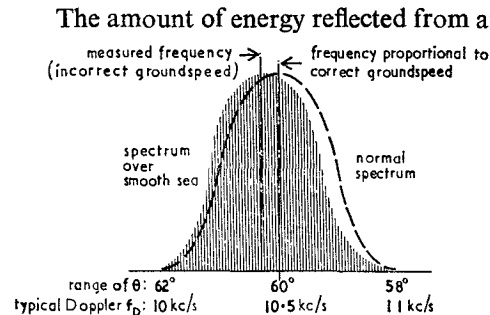


FIG 17. EFFECT OF SEA BIAS ON DOPPLER SPECTRUM

borne Dopplers sea bias effect is reduced by using a land-sea switch. Operation of the switch alters the calibration of the Doppler frequency tracker to provide correct groundspeed measurements over either land or sea.

Aerial Systems

The most effective aerial system for airborne Dopplers is the slotted waveguide linear array (see p330). In a typical four-beam system each beam is directed about 20° from the fore-and-aft axis of the aircraft (port or starboard) and, at the same time, it is depressed towards the ground by an angle of about 60° . In some systems all four beams are operative together; in other systems the beams are switched in turn; and in others the switching is by beam-pairs—one forward and one backward together.

Each beam may be produced by a *linear array* formed by a length of waveguide with a number of inclined slots cut in one narrow wall. Two types of slotted waveguide arrays may be used: an 'anti-phase' array, in which adjacent slots are inclined in *opposite* directions (Fig 18a); and an 'in-phase' array in which all the slots are inclined in the *same* direction (Fig 18b).

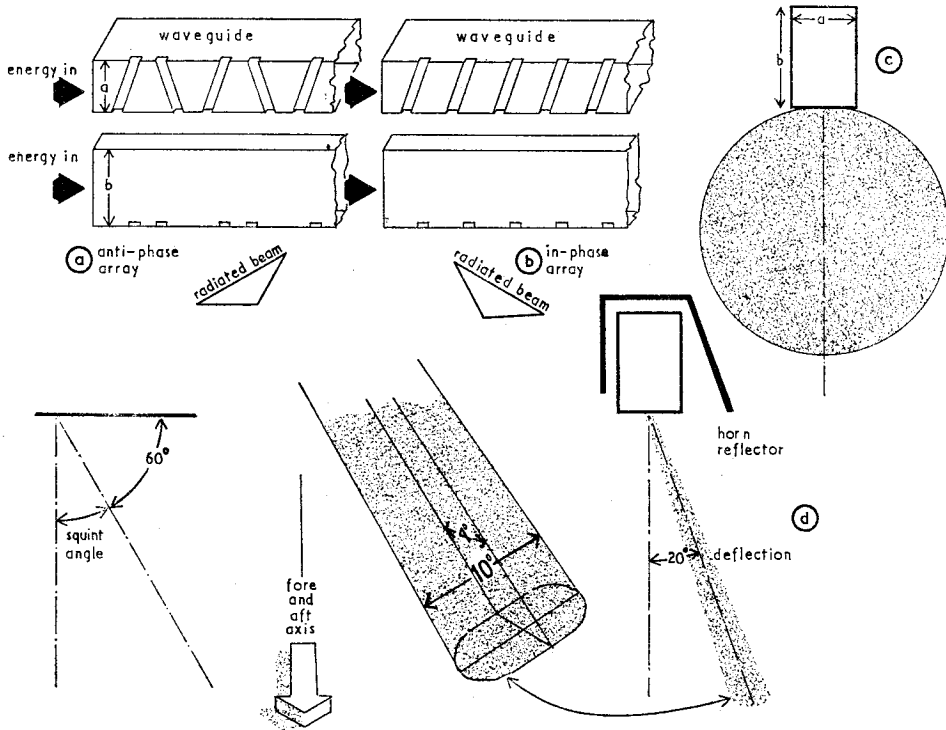


FIG 18. PRODUCTION OF DOPPLER BEAMS BY SLOTTED WAVEGUIDE LINEAR ARRAY

We saw on p331 that, by suitable spacing of the slots in an 'anti-phase' array, the radiation is at some angle to the normal to the array. This angle (the *squant angle*) depends upon the transmitted frequency, the waveguide dimensions, and the slot spacing. By suitable design the squint angle can be made such that we obtain the required depression angle of about 60° .

Similar remarks apply to the 'in-phase' array of Fig 18b. Here, however, because of the in-phase coupling between the slots, the radiation angle is different. If we assume that the 'anti-phase' array radiates in a *forward* direction, then the 'in-phase' array will radiate in a *backward* direction. This is illustrated in Fig 18.

The end view of the slotted waveguide array (Fig 18c) shows that the radiation pattern is approximately circular in this plane. To direct the beam to port or starboard, the waveguide is placed within a horn-type reflector (Fig 18d).

To produce the four beams of the Janus system, four radiators are required—two 'in-phase' and two 'anti-phase' arrays. The aerial assembly is positioned so that the axis of the aerial lies along the fore-and-aft axis of the aircraft. A typical aerial of this type, with its horn-type reflectors, is illustrated in Fig 19 overleaf. To provide beam-pair switching, the transmitter output is applied first to radiators 1 and 3, and then to radiators 2 and 4. The input to the receiver is similarly switched. The switch may be mechanical, ferrite, or semiconductor.

If we apply the Doppler transmitter output first to one end of a slotted waveguide array and then to the other, it may be seen that the radiation angle in the depression plane is *switched* backwards and forwards. It is possible, therefore, to reduce the number of arrays to two, one

'in-phase' and one 'anti-phase', and switch the transmitter feed to alternate ends of the array to produce the usual beam-pair switching. This method is in use in one current airborne Doppler equipment.

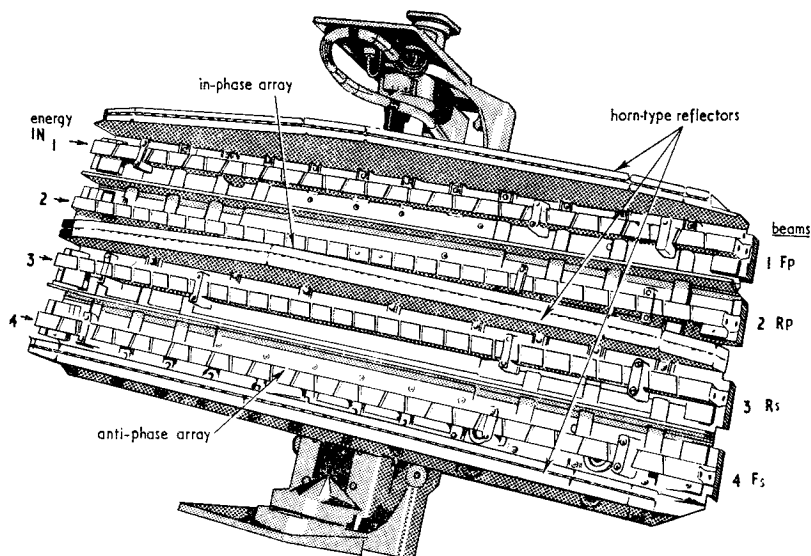


FIG 19. TYPICAL JANUS FOUR-BEAM AERIAL

Frequency Tracking

We saw earlier that the echo signal received in an airborne Doppler equipment contains a *spectrum* of Doppler frequencies. The job of a frequency tracker is to measure the *mid-frequency* of this spectrum in order to determine ground-speed, and to 'follow' the mid-frequency variation as the groundspeed changes. By comparing the signals received from port and starboard aerial beams the drift angle may also be measured. In addition, the frequency tracker usually includes circuits for automatic gain control of the receiver, and for 'memory' operation during periods of inadequate signal strength.

The frequency tracker usually contains a *comparator* circuit in which the detected Doppler beat frequency is compared with the frequency of a *reference source* (Fig 20). The output of the comparator represents the 'error' signal and this is used to alter the frequency of the reference source until it is the same as that of the Doppler input. This is the normal 'null-operated' arrangement. The tuning control of the reference source can then be calibrated in terms of groundspeed.

A convenient reference source that has been much used is the *phonic wheel oscillator* driven by a servomotor. Its frequency can be easily controlled; its output is in the audio range, corresponding to the airborne Doppler range of 2 kc/s to 25 kc/s; and the servomotor speed is a direct analogue of groundspeed. The *distance flown* is given by the *total number of revolutions* of the motor shaft and this may also be displayed on a simple counter.

The phonic wheel oscillator has not previously been discussed in these notes. It is such a common device that it may be helpful to consider its basic principles of operation here.

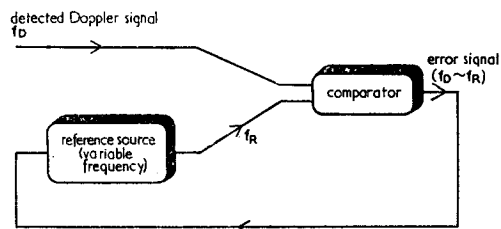


FIG 20. BASIC IDEA OF FREQUENCY TRACKING

Phonic Wheel Oscillators

A phonic wheel oscillator is basically a small electric motor with a toothed wheel mounted on its shaft (Fig 21a). Two pick-up heads, each consisting of a coil wound round a permanent

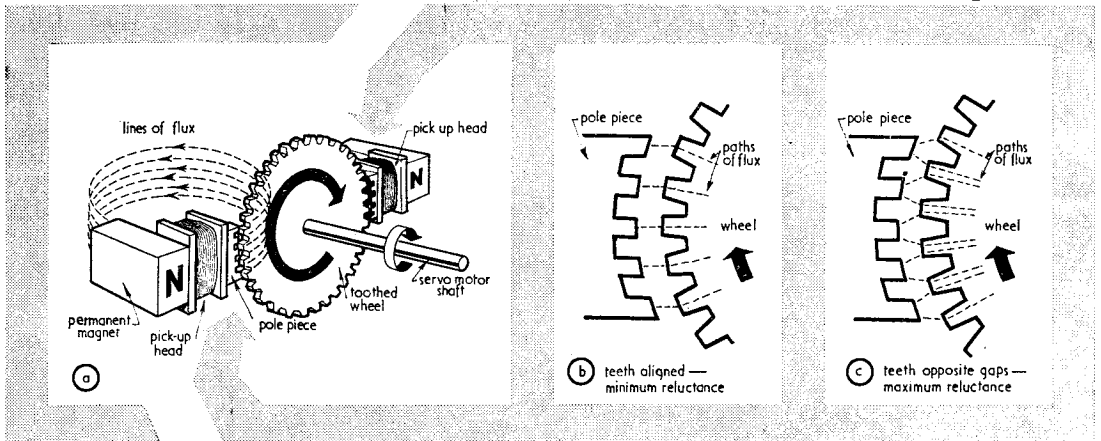


FIG 21. ELEMENTS OF AN ELECTROMAGNETIC PHONIC WHEEL OSCILLATOR

magnet pole piece, are mounted close to the wheel, diametrically opposite each other. Each magnet pole piece has a number of teeth cut in it also, of pitch equal to that of the teeth cut in the wheel. The lines of magnetic flux pass from the pole piece, across the small air gap, to the wheel, thence through air again to the other pole of the magnet.

The flux in the pole pieces depends upon the *reluctance* of the magnetic path. As the wheel rotates, its teeth pass across the teeth on the face of the pole piece. Thus, the reluctance varies from a minimum value (Fig 21b) to a maximum value (Fig 21c). This causes the flux in the pole pieces to vary in a regular manner, the changing flux inducing a voltage in the pick-up coils. One cycle of flux variation takes place during the passage of a wheel tooth from alignment with one pole piece tooth to alignment with the next. Hence, the *frequency* of the induced voltage depends upon the speed of rotation of the wheel and also upon the *number of teeth* cut in the wheel. The two pick-up heads are wired in series to give an alternating current output of a frequency that can be closely controlled by the speed of the servomotor rotating the wheel.

For some applications, *two* reference sources are required. The frequency produced by each source is required to be different, but between the two frequencies we require a *constant percentage difference*, no matter how the frequencies themselves may vary. This may be obtained by mounting *two phonic wheels* on a *common* servomotor shaft and arranging that the numbers of teeth cut on each wheel differ by the same percentage as that required for the two frequencies.

Twin-line Tracking System

The basic outline of a frequency tracker that uses a twin-phonic wheel oscillator source is illustrated in Fig 22.

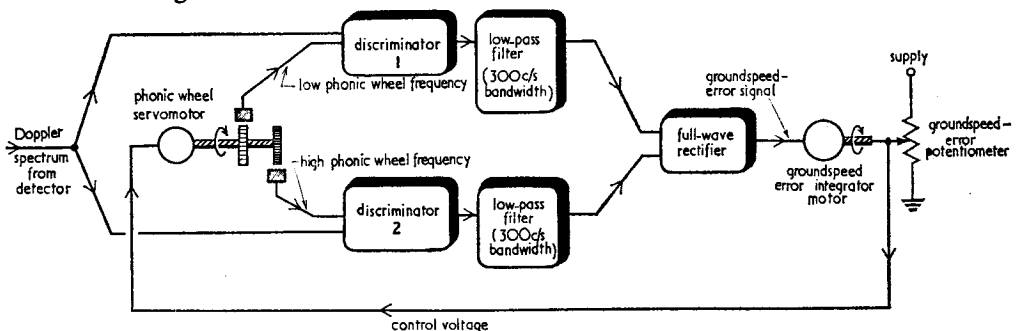


FIG 22. TWIN-LINE FREQUENCY TRACKING SYSTEM, BASIC OUTLINE

Measurement of groundspeed. The two frequencies produced by the phonic wheels differ by about 10% of the mean phonic wheel frequency at any given time and are variable within the Doppler range of frequencies from about 2 kc/s to 25 kc/s. The phonic wheel frequencies are each compared with the Doppler spectrum in separate mixer (discriminator) circuits, the outputs of which drive a servomechanism to adjust the phonic wheel frequencies until they 'straddle' the mid-frequency of the Doppler spectrum. At this point balance has been achieved and the speed of the phonic wheel motor is then a measure of aircraft groundspeed.

The received ground echo, containing the Doppler spectrum, is processed in the receiver, and the detected Doppler spectrum is then applied to two discriminator circuits where it is mixed with the phonic wheel outputs. Discriminator 1 is fed with the low phonic wheel frequency and discriminator 2 with the high frequency. Each discriminator circuit contains a low-pass filter which allows through only those frequencies contained in a band of ± 150 c/s centred on the appropriate phonic wheel frequency. In other words, each discriminator-filter combination acts as a 'gate', each gate passing that part of the Doppler spectrum which overlaps the appropriate phonic wheel frequency by ± 150 c/s. Each band is referred to as a 'window'.

The output from each discriminator is applied to a full-wave rectifier, the amplitude of each discriminator output depending upon the amount of Doppler spectrum contained within the appropriate window.

If the discriminator outputs are equal, as in Fig 23a, the rectifier provides no output and the servo system is static. The phonic wheel motor is now running at a speed proportional to aircraft groundspeed.

If the groundspeed *increases*, the Doppler spectrum shifts to a higher frequency band and the phonic wheel windows are no longer balanced about the mid-frequency of the Doppler spectrum (Fig 23b). The output from discriminator 2 is now *greater* than that from 1, and we have a net output from the full-wave rectifier. This output causes the groundspeed-error integrator motor to rotate, altering the setting of the groundspeed-error potentiometer. This, in turn, alters the input to the phonic wheel drive motor to adjust its speed and hence the phonic wheel frequencies. The *polarity* of the output from the full-wave rectifier is such that the sense of the integrator motor movement *increases* the speed of the phonic wheel motor. The phonic wheel frequencies therefore increase, in step, and when they are once again balanced about the mid-frequency of the Doppler spectrum the integrator motor stops. The speed of the phonic wheel motor is now proportional to the increased groundspeed.

The result of a *decrease* in groundspeed is shown in Fig 23c. The servo system is again activated (in the opposite direction this time) and the phonic wheel motor speed is adjusted until balance is once more achieved.

The speed of the phonic wheel motor (proportional to groundspeed) depends upon the setting of the groundspeed-error potentiometer.

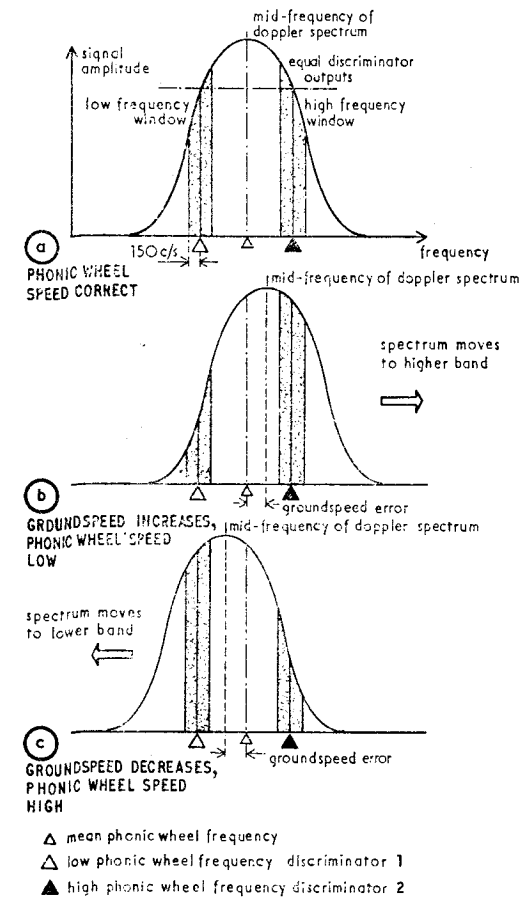


FIG 23. MEASUREMENT OF GROUNDSPD IN TWIN-LINE TRACKING SYSTEM

Hence the setting of the potentiometer shaft may be used as a *direct analogue* of groundspeed. This may be achieved by mounting a torque synchro transmitter on the potentiometer shaft and transmitting shaft angle information to a torque synchro receiver calibrated in terms of groundspeed. (See p538 of AP3302 Part 1B for details of synchros).

Since the speed of the phonic wheel motor at balance is proportional to groundspeed, it may be seen that the *total number of revolutions* can be used as a measure of the *total distance flown*. By gearing a suitable counter to the phonic wheel motor shaft this information may be shown as 'nautical miles flown' (similar to the odometer in a motor car).

Drift Measurement. When a moving-aerial, four-beam system is used in an airborne Doppler, the discriminator circuits used to measure groundspeed may also be used to align the aerial with the aircraft track. If the aircraft is drifting and the aerial is *not aligned* along the track, different Doppler frequencies are obtained as the transmissions are switched from one beam-pair to the next (see p 483). This causes the Doppler spectrum to move up and down the frequency scale in synchronism with the aerial switching (Fig 24). The discriminator outputs are switched in step with the aerial beam switching with the result that, under conditions of aerial non-alignment, the amplitude of the output from one discriminator channel is *consistently greater* than that from the other.

By applying the switched discriminator outputs to a separate full-wave rectifier an output is obtained and this is used to drive another servomotor—the *drift motor*. The polarity of the rectified output will depend upon which discriminator provides the greater output, i.e. upon the *sense* of the drift. The drift motor is thus driven one way or the other and moves the aerial array in azimuth in the correct sense until balance is achieved, whereupon the drift motor stops. The aerial is then aligned along the aircraft *track* and the angle through which it has turned is the drift angle.

It is usual to repeat the drift angle information from the drift motor shaft through a synchro system to a suitably calibrated indicator. For some applications a control differential synchro system is used: this combines the drift angle reading and the heading input from the aircraft compass to give the aircraft track for application to the ground position indicator (GPI).

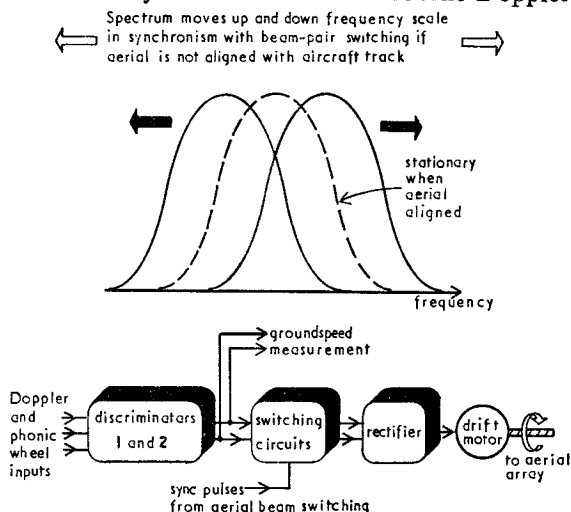
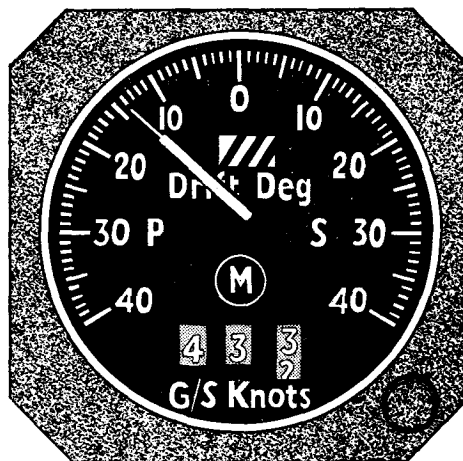


FIG 24. MEASUREMENT OF DRIFT IN TWIN-LINE TRACKING SYSTEM

AGC and 'memory' operation. The a.g.c. voltage in a Doppler system is usually derived from the output of the full-wave rectifier in the frequency tracker and is applied in the normal way to several of the i.f. stages in the receiver. It is also applied to a 'memory' unit. If the signal-to-noise ratio falls to such a level that the Doppler readings tend to become unreliable, the fall in a.g.c. voltage triggers the memory unit. This, in turn, operates relays which break the groundspeed-error and the drift servocircuits; the groundspeed and drift indicators therefore continue to record the *last known readings*. In addition 'distance gone' information will continue to change at its *last known rate*. A flag marked M is automatically shown on the indicators when the equipment is operating on memory. As soon as the signal-to-noise ratio rises sufficiently the equipment reverts to normal operation and the M flag is withdrawn.

Indicators. Airborne Dopplers are fitted with at least two indicators, both mounted on the cockpit instrument panel. One instrument, the groundspeed and drift indicator, combines a three-digit odometer-type counter for presenting groundspeed and a pointer for showing drift angle. The drift angle reading is obtained by direct torque synchro action from the drift motor shaft. Similarly, the torque synchro groundspeed output from the groundspeed-error integrator motor shaft is used to feed a control synchro transformer. The output from this drives a small servomotor which adjusts the groundspeed counters in the indicator. The groundspeed (GS) range normally available is 0 to 999 knots and the drift angle is calibrated up to a maximum of 40° port and starboard. Memory (M) and on-off (striped) flags are also incorporated on the instrument face (Fig 25a).



The other instrument is the distance-flown indicator (Fig 25b). This is a five-digit odometer-type counter which indicates total distance flown up to 99,999 nautical miles in steps of one nautical mile. Other ranges are available. The input to this unit is from the torque synchro on the phonic wheel motor shaft, and the torque synchro receiver is geared to the counter to provide the indication. A reset control is available to return the reading to zero when required.

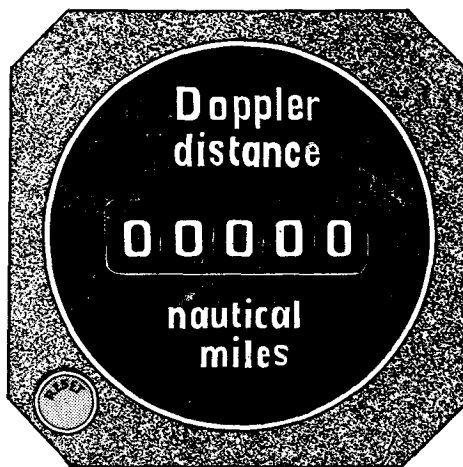


FIG 25. TYPICAL AIRBORNE DOPPLER INDICATORS

Summary. The elements of a twin-line phonic wheel frequency tracker incorporating all the points discussed so far are illustrated in Fig 26. The servomotors normally include tachogenerators to provide velocity feedback damping, thereby reducing 'hunting'. The feedback elements have been omitted for clarity, but information on servomechanisms is given in p545 of AP 3302 Part 1B.

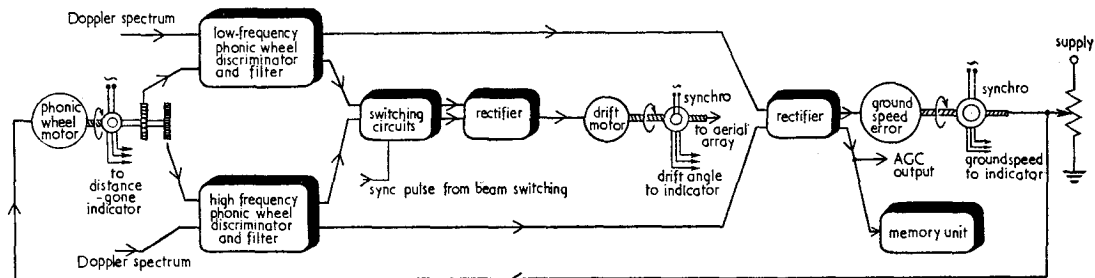


FIG 26. SIMPLIFIED BLOCK DIAGRAM OF TWIN-LINE PHONIC WHEEL FREQUENCY TRACKER

Single-line Tracking

The single-frequency (or single-line) tracking system continuously aligns a *single* phonic wheel frequency with the *mean frequency* of the Doppler spectrum. The Doppler spectrum is applied, as before, to two discriminator circuits, each of which has also an input at the phonic wheel frequency (Fig 27). The *same* phonic wheel frequency is applied to both discriminators

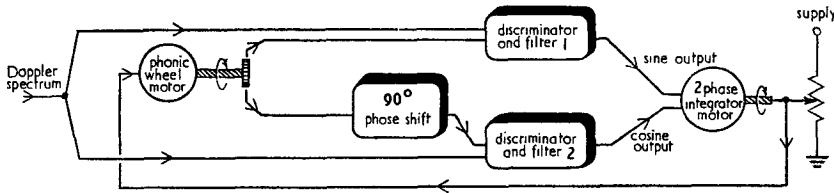


FIG 27. BASIC SINGLE-LINE TRACKING SYSTEM

but with a 90° difference in phase. Thus, one discriminator produces a 'sine' output and the other a 'cosine' output. For this reason single-line trackers are often called '*sin-cos*' trackers.

The discriminator outputs are applied to a two-phase integrator motor. The effect of mixing is that if the phonic wheel frequency is *higher* than the mean frequency of the Doppler spectrum the integrator motor will turn in one direction; if the phonic wheel frequency is *lower* the motor turns in the other direction.

The integrator motor drives a potentiometer which controls the voltage driving the phonic wheel motor. Movement of the integrator motor will therefore cause the phonic wheel motor to speed up or slow down, thus changing the phonic wheel frequency. When the phonic wheel frequency is aligned with the mean frequency of the Doppler spectrum the integrator motor stops and the phonic wheel motor then runs at a steady speed—proportional to aircraft ground-speed.

Electronic Tracking Systems

The trend towards solid state devices, with their increased reliability, has meant that modern Doppler equipments are tending to use single-line '*sin-cos*' trackers with an *electronic* oscillator, controlled by an *electronic* integrator, producing the local signal to be applied to the discriminators. In addition, with the advent of c.w. Dopplers and no beam switching, it is usual to track each beam *separately*. Each tracker (one for each beam) then produces an output proportional to aircraft velocity *in the direction of the beam*, and by feeding these outputs to a computer the forward, sideways, and vertical velocity components can be resolved to provide groundspeed and drift.

The simplified block diagram of a three-beam, fixed-aerial, c.w. Doppler is illustrated in Fig 28.

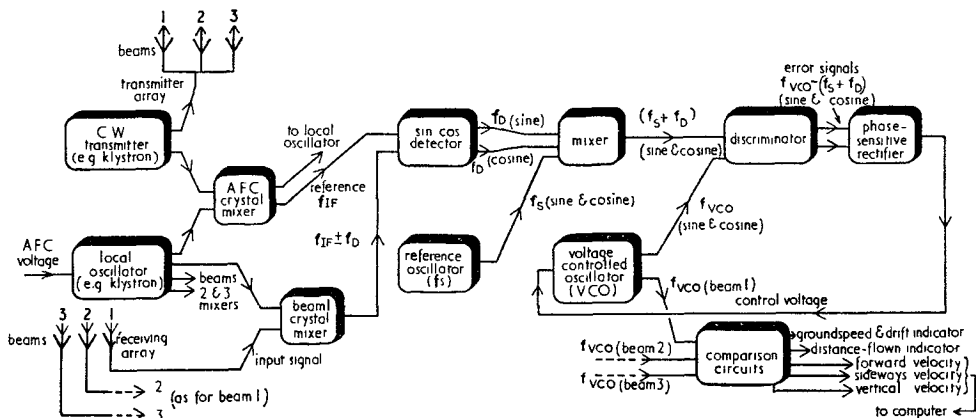


FIG 28. SINGLE-LINE TRACKING IN A THREE-BEAM CW DOPPLER, BASIC OUTLINE

Returned signals, containing the Doppler spectrum corresponding to aircraft velocity in the direction of the appropriate beam, are processed by the receiver in three separate channels. Each of the three Doppler echoes is mixed with the local oscillator output to produce separate i.f. outputs containing the Doppler spectrum ($f_{IF} \pm f_D$). In Fig 28 the arrangement for beam 1 only is shown. The mixer output at $f_{IF} \pm f_D$ is further mixed with the reference signal f_{IF} in a sin-cos detector to produce two Doppler outputs at f_D , phase-shifted 90° to each other. This pair of signals is then mixed with locally-generated sin-cos signals of frequency f_s to produce sin-cos outputs of $f_s + f_D$ (or $f_s - f_D$ in a backward beam). The signals at $f_s + f_D$ are then applied to a discriminator along with the sin-cos outputs of a *voltage-controlled oscillator* (v.c.o.). The resulting output from the discriminator consists of two equal 'error' frequencies of $f_{vco} \sim (f_s + f_D)$, 90° phase-shifted to each other, and these are applied to a phase-sensitive rectifier.

The *sign* of the phase-sensitive rectifier output depends upon whether f_{vco} is above or below the mid-frequency of the Doppler spectrum signal. The phase-sensitive rectifier output is used to apply a control voltage to the v.c.o. to achieve balance, f_{vco} then being a measure of the aircraft velocity in the direction of the beam. In addition, f_{vco} will automatically *follow any variation* in the Doppler spectrum caused by aircraft movement.

Outputs from the three voltage-controlled oscillators (one for each beam) are fed to comparison circuits where pulse outputs proportional to forward, sideways, and vertical velocities are obtained. These are available for application to a navigation computer. In addition, by combining appropriate inputs we can also obtain outputs proportional to groundspeed, drift angle, and distance flown, for application to the indicators.

Computers for Airborne Dopplers

Many different types of navigation computers are available for use with airborne Dopplers. Such computers are normally considered as part of the instrument system of an aircraft and their general principles of operation are, therefore, dealt with in AP3302 Part 4. Here, we wish to round off the general principles of airborne Doppler and, to complete the picture, it is worth considering the elementary outline of a typical computer system (Fig 29).

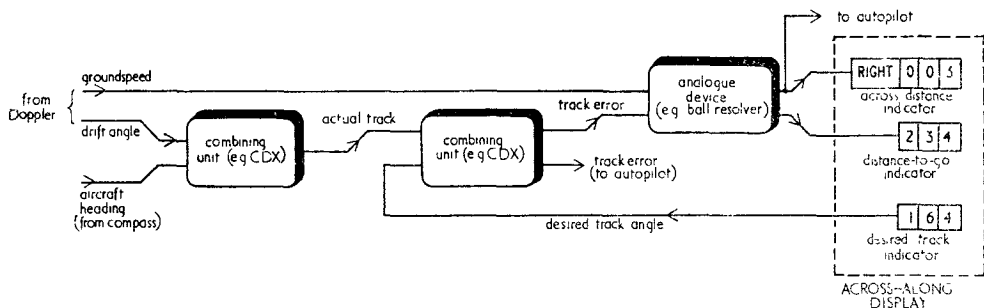


FIG 29. OUTLINE OF ONE TYPE OF DOPPLER COMPUTER SYSTEM

In the arrangement of Fig 29 the Doppler equipment applies two inputs to the computer: an input proportional to groundspeed, and one proportional to drift angle. The drift angle and the aircraft heading (from the aircraft compass) are applied to a combining unit, which may be a

control differential synchro. The output from the synchro then represents the *actual track* of the aircraft. The track angle, in turn, is compared with the *desired track* angle, obtained from the pre-arranged flight plan, and from these two inputs the *track error angle* is obtained. The track error angle may be fed to the autopilot (which proceeds to correct the aircraft track) and it may also be applied to an analogue device known as a 'ball resolver' (see AP3302 Part 4). The Doppler output proportional to groundspeed is also applied to the ball resolver which then provides outputs in terms of *distance along* and *distance across* the desired track.

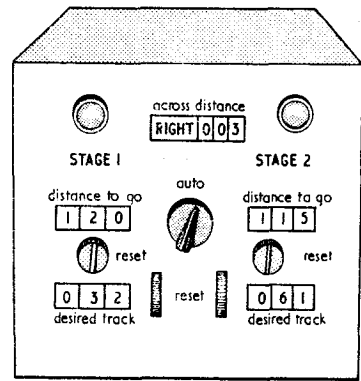
A typical along-across display is illustrated in Fig 30a. This type of display is particularly useful when the pre-arranged flight plan requires the aircraft to fly several 'legs' in its journey from X to Y (Fig 30b).

The stage 1 part of the indicator is set up by the manual controls to satisfy the requirements of the first leg, and stage 2 is similarly set up for the second leg. If the system is switched to 'automatic' then on take-off the stage 1 part of the display is operative. The distance-to-go reading decreases and any deviation from the desired track is indicated in the across-distance counter. With about 10 miles to go on the first leg, the stage 1 warning lamp (which has been glowing steadily) starts to flash, and when the first leg is completed the equipment switches to stage 2. Whilst stage 2 is operative, stage 1 can be reset for the *third leg* of the flight. This procedure is repeated as required.

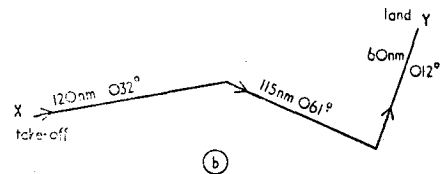
Other types of display are available for operation from the computer.

These include:

a. Roller map (Fig 31a). This gives a continuous indication of the position of the aircraft on a suitable map. Aircraft movement along track is shown by movement of the map from top to bottom of the instrument. Position along track is shown by a cursor and movement *across* track is shown by lateral displacement of the cursor. A pen which moves with the cursor may be fitted.

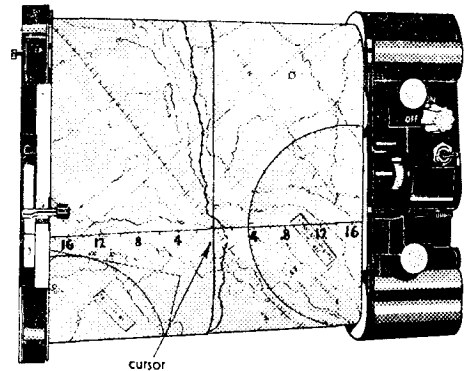


(a)



(b)

FIG 30. SIMPLE FLIGHT PLAN AND USE OF ALONG-ACROSS DISPLAY



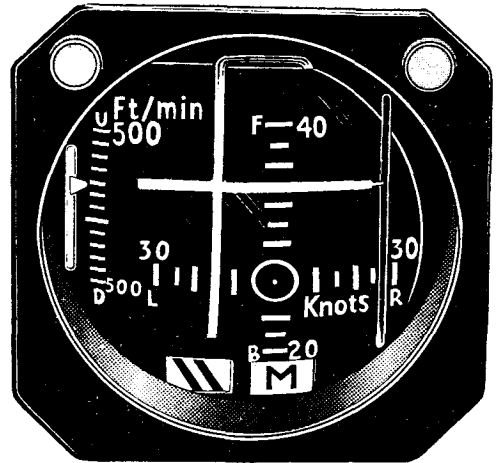
(a)

FIG 31.

b. *Hover meter* (Fig 31b). This is used in helicopters where movement may be in any direction. The three components of aircraft velocity (forward, sideways, and vertical) are fed to the meter as inputs. Typical ranges for meter readings are:

- (1) *Along-heading velocity*. From - 20 knots (backwards) to + 40 knots (forward).
- (2) *Across-heading velocity*. Approximately ± 30 knots (left-right).
- (3) *Vertical velocity*. In the range ± 500 feet per minute (up or down).

In Fig 31b the readings indicated are + 25 knots (forward), - 15 knots (sideways, left), and + 200 ft/min vertical, upwards).



(b)
FIG 31.

Conclusion

In this chapter we have been concerned only with the basic principles of airborne Doppler and the application of these principles, in a general way, to typical systems. Equipments in service vary considerably in circuit detail and in their application of the described principles. Thus, when carrying out servicing or maintenance tasks on an actual equipment it will be necessary to consult the official Air Publication for the equipment.

In common with other electronic devices, airborne Dopplers are being developed which take advantage of microelectronic and integrated circuit techniques. It is expected that this will lead to an improvement in reliability as well as reducing the overall size and weight of the equipment. New circuit techniques will obviously emerge as a result of this development but the basic principles, as described in this chapter, remain the same.

MOVING-TARGET INDICATION [MTI] RADAR

Introduction

We have mentioned earlier in these notes that it is possible for a skilled operator, using a normal pulsed radar, to discriminate between fixed objects and moving targets. He does this by noting the *change* in moving-target position on the p.p.i. from scan to scan. However, this is a rather indirect means of distinguishing between moving targets and fixed objects. If we are using a large p.p.i. range-scale, the blip representing the moving target will change position only *very slowly* with time and several scans may be required to indicate a moving target. A more satisfactory method of discriminating between fixed objects and moving targets is to use a combination of pulse and Doppler techniques in *moving-target indication (m.t.i.) systems*.

The ability to discriminate between fixed objects and moving targets is not the only advantage of the m.t.i. radar. Its main advantage is its ability to reduce returns from fixed objects to such an extent that moving targets can be 'seen' in conditions of severe clutter (Fig 1). In practice, an m.t.i. radar can extract the moving-target echo from the clutter even if the fixed-clutter echo is 30 db (1000 times) greater than the moving-target echo.

In earlier chapters we discussed other ways of reducing clutter (p427). Some of these methods are more effective than m.t.i. in reducing *weather clutter* (e.g. the use of circular polarization). But m.t.i. systems are the most effective means of reducing clutter caused by *stationary objects* (permanent echoes).

In m.t.i. systems, the radar information is delayed or stored and then used to cancel the information obtained a short interval of time later. In this way, echoes which *have not changed with time* are cancelled (stationary objects), whilst echoes which *have changed with time* are not cancelled (moving targets.)

Two basic forms of m.t.i. are used. The first compares the returned echoes in *successive scans*, and is known as *area m.t.i.* We know that the permanent echo or clutter pattern on a p.p.i. remains unchanged and is repeated from scan to scan. If we could arrange to compare the picture on one scan (Fig 2a) with that on the succeeding one (Fig 2b) the clutter patterns could be subtracted to give complete cancellation. Only those echoes which have *changed position* from one scan to the next (i.e. moving targets) are now displayed, as shown in Fig 2c. Area m.t.i. has limited use in the Service. Because of this it will not be considered further in these notes.

The second form of m.t.i. compares the returned echoes in successive recurrence periods (*pulse-to-pulse comparison*). The pulse-to-pulse comparison method of m.t.i. is more common and gives the better results. We shall now consider it in more detail.

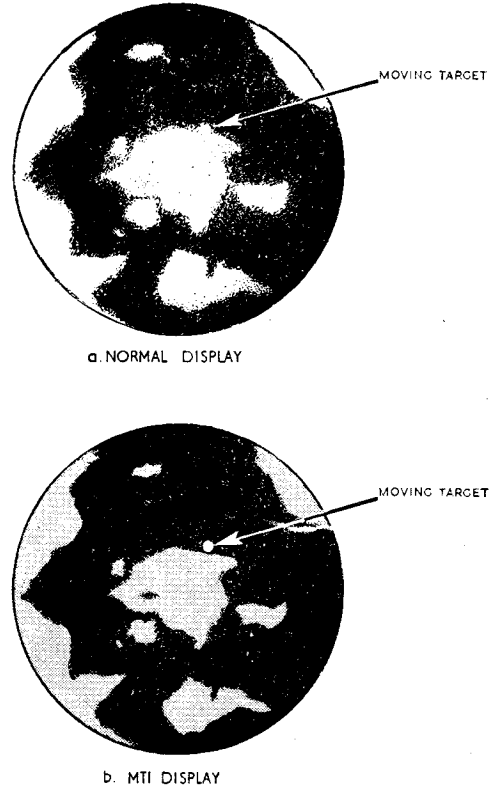


FIG 1. MOVING TARGET IN SEVERE CLUTTER

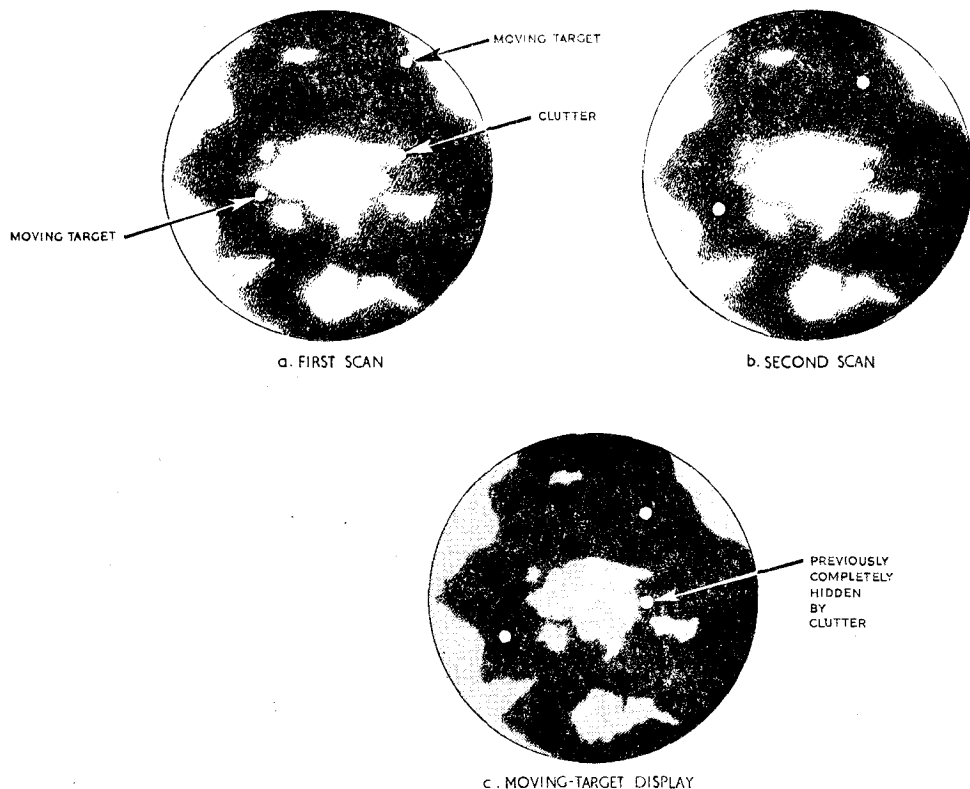


FIG 2. PRINCIPLE OF AREA MTI

Principle of MTI Using Pulse-to-Pulse Comparison

We know from a previous chapter (see p446) that the *frequency* of the echoes from a fixed object is the *same* as that of the transmitted signal. On the other hand, the frequency of the returns from a *moving target* differs from that of the transmitted signal by an amount that depends upon the relative velocity of the target. This difference frequency is the *Doppler shift* and it occurs whether the transmitter output is c.w. or pulsed. The presence of a Doppler shift in either case indicates a moving target.

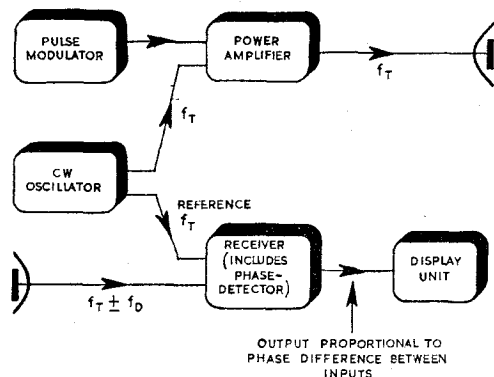


FIG 3. SIMPLE MTI RADAR

mitter are controlled. We shall see later that, in this way, the oscillator provides a *reference signal* against which the phase of the echo pulse can be compared in the receiver.

A normal radar receiver contains r.f. and i.f. stages in which the received echo signals are

amplified and converted to video. The m.t.i. receiver contains similar circuits. However, in the m.t.i. receiver, pulse-to-pulse target motion is detected by a phase-comparison method, so that we shall also find *phasing circuits* in the i.f. stages of the receiver.

The relative phase of the reference signal from the c.w. oscillator and the target echo signal are compared in a *phase-sensitive detector* in the receiver (see page 450). For stationary objects, the Doppler shift f_D is zero and the phase detector provides a *constant-amplitude output* whose magnitude depends upon the *constant* phase difference between the two inputs. However, when the target is in motion relative to the radar, f_D has a value other than zero and the resulting phase difference between the reference signal and the target echo signal is *continuously changing* because of the frequency difference between the two inputs. The output from the phase detector now varies.

The result of comparing the relative phase of the reference signal and a moving-target echo signal is shown in Fig 4. The radiated pulses are illustrated in Fig 4a and the resulting video output from the receiver is shown in Fig 4b. This assumes that the Doppler shift f_D is small in relation to the reciprocal of the pulse duration—the usual case for aircraft targets. Fig 4b shows that when the target is in motion the video output pulses vary in both amplitude and polarity within an envelope of the Doppler shift. Several pulses are required for satisfactory extraction of the Doppler signal. Thus several target 'hits' per scan are necessary in m.t.i. radars. The p.r.f. and the scanning rate must be adjusted accordingly. The waveform illustrated in Fig 4b is referred to as *bipolar video* because the pulses contain both positive and negative amplitudes.

If the target has a high relative velocity, such as may occur with a ballistic missile or a satellite, the Doppler shift f_D can be *much larger* than the reciprocal of the pulse duration. When this applies, the Doppler signal can be extracted from a *single* pulse.

Moving-target Indication on a Type A Display

If we apply the bipolar video output of an m.t.i. receiver to a type A display the result will be as shown in Fig 5. Stationary objects contain no Doppler shift and the amplitude of fixed-object echoes remains the same from pulse to pulse. Moving-target echoes produce a bipolar video output which varies in amplitude and polarity from pulse to pulse at a rate corresponding to the Doppler shift. Thus, during successive pulse repetition intervals 1 to 5 the picture on the type A display will change as shown. The result of superimposing these successive sweeps on the type A display is illustrated by the bottom waveform. This is the picture which is actually seen. Note again that several sweeps (pulse repetition intervals) are required to differentiate between fixed objects and moving targets. During the time of successive sweeps the moving targets produce a 'butterfly' effect on the type A display.

Moving-target Indication on a PPI Display

The butterfly effect cannot be used with a p.p.i. display because variations in amplitude of the moving-target echoes are not readily apparent on a p.p.i. Some other method of moving-target indication must therefore be used.

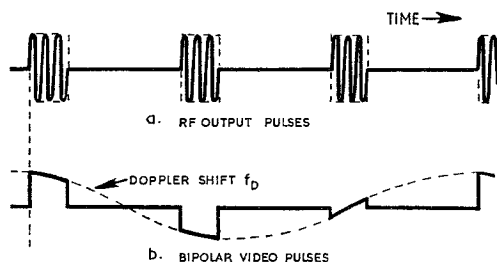


FIG 4. EFFECT OF MOVING TARGET ON MTI RECEIVER VIDEO OUTPUT

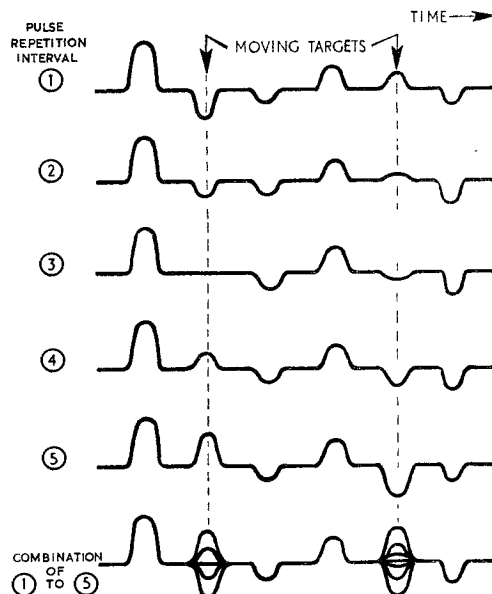


FIG 5. MOVING TARGET INDICATION ON A TYPE 'A' DISPLAY

Fig 5 shows that the permanent echoes (clutter) remain the same from one sweep to the next, whilst the echo from a moving target varies in amplitude and polarity. Thus if we could arrange to subtract sweep 1 from sweep 2, sweep 2 from sweep 3, and so on, the permanent echoes would disappear and only the moving-target echo would remain. This is the principle adopted in most m.t.i. systems.

To achieve this pulse-by-pulse subtraction the video portion of the receiver is divided into two channels. One is a normal video channel. In the other, the bipolar video output of the phase-sensitive detector is *delayed* by one pulse repetition period (equal to the reciprocal of the p.r.f.). The outputs from the two channels are then subtracted from one another. In this way the permanent echoes are removed, whilst subtraction of the moving-target echoes of varying amplitudes results in an *uncancelled residue*.

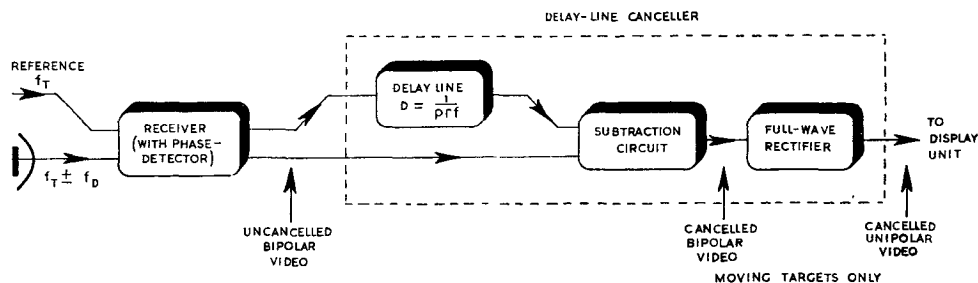


FIG 6. CANCELLATION OF FIXED-OBJECT ECHOES IN DELAY-LINE CANCELLER

The output of the subtraction circuit is again bipolar video. But this time it contains only *moving-target* information. For intensity-modulation of a p.p.i. display we require *unipolar video*, i.e. voltages all of the same sign. Thus, a full-wave rectifier is needed to convert the bipolar output of the subtractor circuit into unipolar video. A schematic outline of an m.t.i. receiver with a *delay-line canceller* is shown in Fig 6. The delay line used in the canceller can take one of several forms. We shall discuss typical delay lines later.

MO-PA Type of MTI Radar

So that the phase-sensitive detector in the i.f. stages of the receiver can recognise the pulse-to-pulse phase changes in a moving-target echo, the returning signals must be compared in phase with an oscillator whose phase is 'locked' to that of each transmitted pulse. In Fig 7 this 'coherent' reference signal is supplied by an oscillator called the *coho* (*coherent oscillator*).

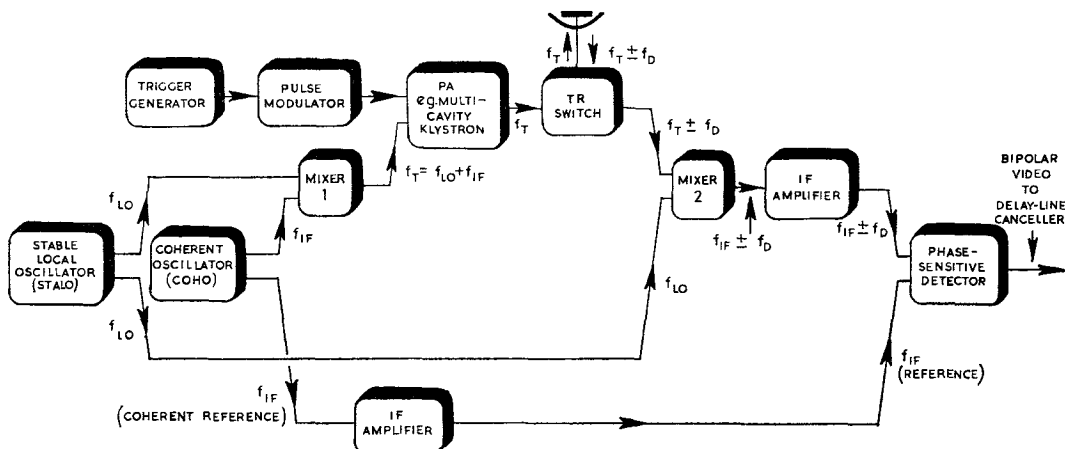


FIG 7. MO-PA TYPE OF MTI RADAR

In this circuit the coho is a stable oscillator (very often crystal-controlled) whose frequency is the same as the i.f. used in the receiver. Besides providing the reference signal for the receiver, the output f_{IF} of the coho is mixed with the local oscillator signal at frequency f_{LO} in mixer 1 to provide the transmission frequency $f_T = f_{LO} + f_{IF}$. The local oscillator must also be stable in frequency and is referred to as the *stalo* (stable local oscillator). The output of mixer 1 at frequency f_T is applied to the p.a. valve (e.g. a multi-cavity klystron). There it is amplified and pulse-modulated to provide the high-power r.f. output pulses at frequency f_T .

The echo signal from a moving target is Doppler-shifted in frequency so that the input to the receiver is at frequency $f_T \pm f_D$, where f_D is the Doppler shift. This input signal is mixed with the stalo signal at frequency f_{LO} in mixer 2 to provide the i.f. signal at $f_{IF} \pm f_D$. The reference signal at f_{IF} from the coho and the i.f. echo signal are then amplified separately and applied as the two inputs to a phase-sensitive detector. The detector produces an output proportional to the *phase difference* between the two inputs. As explained earlier, the output for a fixed-object echo is a constant amplitude voltage. For a moving-target echo, the Doppler shift ensures a continuously-changing phase difference between the two inputs. The detector output is then *bipolar video* containing both fixed-object and moving-target information. This output is applied to the delay-line canceller to remove the fixed-object information.

In Fig 7, the transmitted signal is in phase with the reference signal in the receiver. This is ensured by actually *generating* the transmitted signal from the coho. The stalo is included to provide the necessary frequency conversion from the i.f. to the transmitted r.f. frequency. It may be thought that any variation in the phase of the *stalo* will upset this phase-coherence. However, examination of Fig 7 will show that any variation in stalo phase shift is cancelled on reception because the stalo is *common* to the transmitter and receiver chains. If the phase of a fixed-object echo changes because of a change in the phase of the transmitter output due to the stalo, this phase change is automatically accounted for and cancelled on reception.

We have mentioned that a multi-cavity klystron may be used as the p.a. valve in the transmitter. Other high-power valves may also be used. These include planar triodes or tetrodes, travelling-wave tubes, and amplitrons (see p276).

Magnetron Type of MTI Radar

Not all radar transmitters use an mo-pa system. In many radars the high-power p.a. stage is omitted and the transmitter r.f. output is provided by a power oscillator, such as a *magnetron*. Fig 8 illustrates the block diagram of a magnetron type of m.t.i. radar. In a pulsed oscillator, the phase of the radiated signal is quite *random* from pulse to pulse. Thus the phase of the

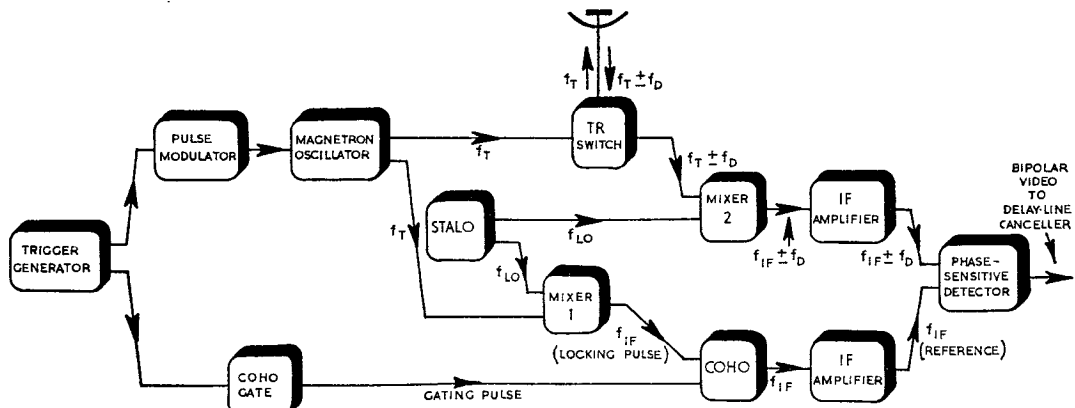


FIG 8. MAGNETRON TYPE OF MTI RADAR

radiated signal cannot be controlled in the way that it is in the m.o.p.a. type of transmitter. Some other means of providing a reference signal against which to compare the phase of the received echo is therefore necessary.

The transmitter portion of the radar is conventional and provides the high-power r.f. output pulses to the aerial at frequency f_T . The phase of the radiated pulses is allowed to be random.

A small fraction of the pulsed transmitter output at f_T is mixed with the stalo output at f_{LO} in mixer 1 to produce an 'i.f. locking pulse.' This locking pulse is applied to the coho and causes the phase of the coho to be re-set every recurrence period so that it locks 'in step' with the phase of the i.f. locking pulse (Fig 9).

The phase of the i.f. locking pulse is the same as that of the transmitted pulse. Hence the phase of the coho is related to the phase of the transmitted pulse and so provides a reference signal for echoes received from *that particular pulse*. When the transmitter fires again, another i.f. locking pulse is generated to re-lock the phase of the coho for that pulse repetition period. The coho is gated from the trigger unit; it cuts on coincident with the transmitted pulse and remains operative, producing the reference i.f. oscillations, until cut off by the gating waveform after a time dependent upon the required range.

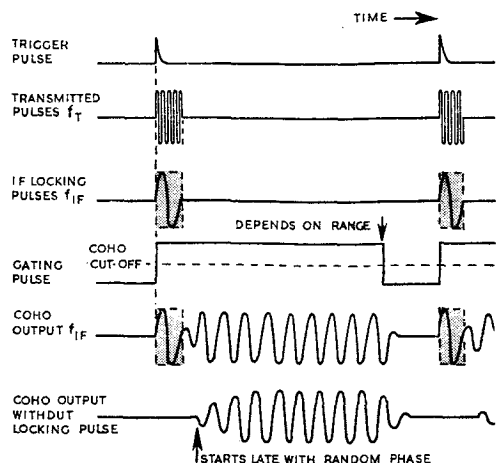


FIG 9. PHASE RELATIONSHIPS IN MAGNETRON TYPE OF MTI RADAR

The received echoes at $f_T \pm f_D$ are also mixed with the stalo output at f_{LO} , and the resulting i.f. signals at $f_{IF} \pm f_D$, after amplification, are applied to a phase-sensitive detector along with the reference i.f. signal from the coho. The output from the detector is then the required bipolar video which is applied to the delay-line canceller.

We have seen that the phase of the transmitted pulses will vary in a random manner from pulse to pulse so that the phase of the echo pulses from fixed objects will vary in the same manner. However, since the transmitted and received pulses are separately mixed with the same local oscillation, the phase of the coho and that of the received i.f. signals will vary *in an identical manner*. This assumes that the stalo is stable in frequency during the period between the transmission of a pulse and the return of the echo. In practice, the stalo is stable to the order of 30 c/s in 3,000 Mc/s. If there is identical pulse-to-pulse phase variation in the coho and echo signals then, for fixed objects, there is *no change* in phase difference between the two inputs to the phase-sensitive detector, and the amplitude of the bipolar video output is *constant*. For a moving target, the echo pulses have an *additional* phase variation because of the Doppler shift and, since this phase variation is *not shared* by the coho, a bipolar video output *varying in amplitude and polarity* in accordance with the Doppler shift is produced.

Delay Lines

The delay line used in the delay-line canceller of Fig 6 is required to introduce a delay equal to the pulse spacing of the radar. We have seen earlier in these notes that the p.r.f. and hence the pulse spacing, depends upon the required range of the radar. For a typical ground-based air-surveillance radar delay times as long as several milliseconds are therefore required. Such relatively long delay times cannot easily be obtained with the LC-type of delay line considered in p367. To achieve the necessary delay we use *acoustic* or *ultrasonic* delay lines.

A simplified ultrasonic delay line is illustrated in Fig 10. The input signal is an r.f. carrier at a frequency of about 20 Mc/s, modulated by the bipolar video output of the phase detector. This signal is converted into mechanical vibrations by a piezo-electric quartz crystal transducer. The vibrations are then transmitted through the delay-line medium to the receiving transducer, which converts the mechanical vibrations back into corresponding electrical variations. The delay introduced by the line depends upon the delay-line medium and upon its length.

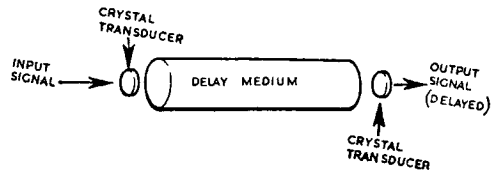


FIG 10. BASIC ULTRASONIC DELAY LINE

One of the earliest forms of ultrasonic delay lines used a cylinder filled with liquid mercury. The transit time of ultrasonic energy in mercury is such that to introduce a delay of 1 millisecond a tube nearly 5 feet long is required. This is an obvious disadvantage.

A more compact package can be obtained by 'folding' the line back on itself one or more times. A mercury delay line using a seven-fold system is illustrated in Fig 11. It is designed to operate at a carrier frequency of 20 Mc/s and to introduce a delay of 1350 microseconds.

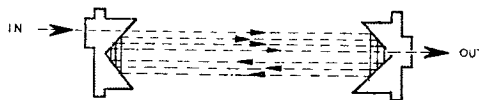
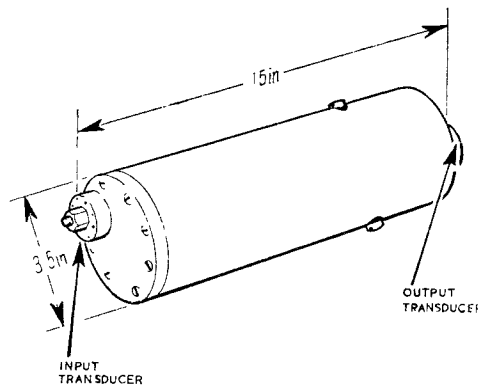


FIG 11. MERCURY DELAY LINE

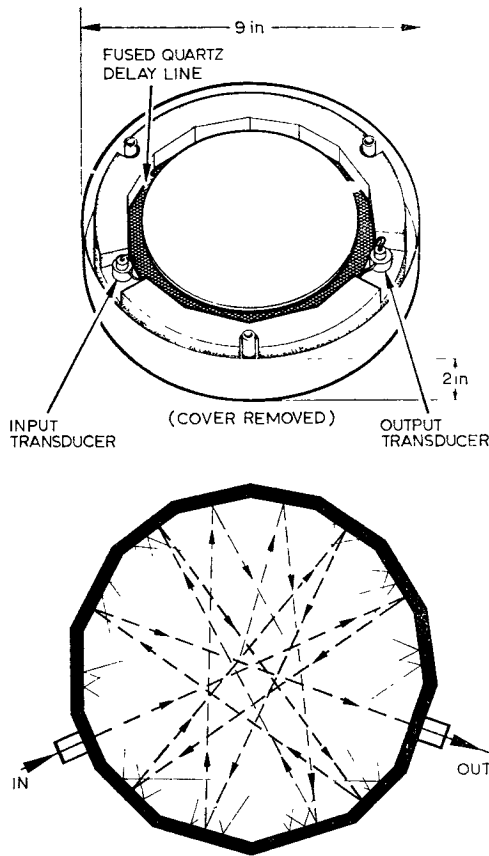


FIG 12. QUARTZ DELAY LINE

Another type of ultrasonic delay line is illustrated in Fig 12. This is a solid delay line of fused quartz. Its shape is an irregular polygon and the piezo-electric crystal transducers are bonded to two of the faces. When the input transducer is excited electrically by the signal, an ultrasonic signal is emitted into the quartz delay line and reflected internally a large number of times before emerging at the receiving transducer. There it is converted back into electromagnetic signal energy. The quartz delay line illustrated in Fig 12 is designed to have 32 internal reflections and a delay time of 1250 microseconds.

The factors which are important in delay lines, apart from the delay time and the operating frequency, include:-

- a. Attenuation introduced by line; the line losses can be quite large in practice.
- b. Bandwidth of line and subsequent distortion.
- c. Range of operating temperatures.
- d. Dimensions and weight of line.
- e. Elimination of unwanted responses.

One type of unwanted response is the 'third-time-round signal'. This is produced by signals being reflected from the receiving transducer back up the line to the transmitting transducer where they are again reflected towards the receiving end. The delay is *three times* that of the original delay. Such responses must be reduced to a minimum.

A comparison of the characteristics of a 1250 microsecond quartz delay line and a mercury delay line, both designed for an operating frequency of 20 Mc/s, is shown in Table 1.

Characteristic	Quartz	Mercury
Insertion loss (attenuation)	60db	95db
Third-time-round response	40db below main signal	50db below main signal
Bandwidth	6 Mc/s	4 Mc/s
Temperature range	— 50°C to + 80°C	— 20°C to + 60°C
Weight	4 lbs	40 lbs
Dimensions	9 inches diameter 2 inches depth	3.5 inches diameter 15 inches long

TABLE 1. CHARACTERISTICS OF DELAY LINES

Examination of this table shows that the fused quartz delay line is better than the mercury delay line in all respects except for the third-time-round response. Most modern m.t.i. delay-line cancellers use the solid quartz delay line.

Delay-line Canceller

The broad outline of the function of a delay-line canceller is given earlier in this Chapter (see p 502). There we showed that the input to the delay-line canceller is the bipolar video from the phase detector in the i.f. stages of the receiver. This input contains both moving-target and fixed-object information. The output from the delay-line canceller is the required *moving-target* information for operation of the display, and the fixed-object information has been suppressed. We must now examine the delay-line canceller in a little more detail to see how these results are obtained.

Fig 13 illustrates the simplified block diagram of one type of delay-line canceller. In this circuit the uncanceled bipolar video output of the phase detector is used to *frequency-modulate* an oscillator operating at a carrier frequency f_1 (typically around 95 Mc/s). After amplification and amplitude-limiting in the line driver, the frequency-modulated carrier at the frequency f_1 is split into two channels, a delayed channel and an undelayed channel.

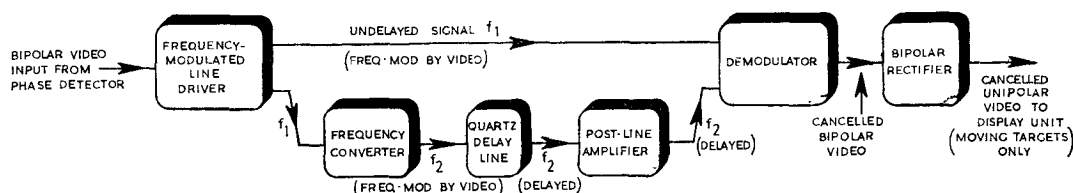


FIG 13. BLOCK DIAGRAM OF SIMPLIFIED DELAY-LINE CANCELLER

The bipolar video signal is not applied directly to the delay-line canceller because the piezo-electric transducers in the delay line of the delayed channel would tend to introduce distortion. By using the video signal to *modulate* an r.f. carrier this difficulty is overcome. The video signal could be used to *amplitude-modulate* the carrier. This is not usually done in practice because, with amplitude modulation, the gains of the delayed and undelayed channels

must be kept in perfect alignment if complete cancellation of fixed-object echoes is to be obtained. In this case the *amplitudes* of the signals in the two channels are being compared and if we have a variation in amplitude of the signal from one channel, due to a difference in gain, an error results. With a *frequency-modulated* input the gain adjustment between the two channels is not critical. Here we are relying on a comparison of *frequency* variations between the two channels.

The undelayed f.m. carrier at frequency f_1 is applied as one input to the block marked 'demodulator'. In the other channel the f_1 carrier is first applied to a frequency-converter. This contains an oscillator, a mixer and an amplifier. The signal from the oscillator is mixed with the input f.m. carrier at f_1 and the resulting output from the mixer is a signal at the difference frequency f_2 (typically 20 Mc/s), frequency-modulated by the bipolar video. The f_2 signal is amplified and applied to the quartz delay line, where it is delayed by a time equal to the radar pulse spacing. After further amplification in the post-line amplifier—inserted to counteract the attenuation of the delay-line—the delayed f.m. carrier at f_2 is applied as the second of the two inputs to the demodulator.

The demodulator block contains a mixer, an amplifier-limiter and a frequency-discriminator (detector). The mixer combines the undelayed f.m. carrier at f_1 and the delayed f.m. carrier at f_2 in such a way that fixed-object echoes on each of the two inputs cancel each other. The output of the mixer is at a difference frequency f_3 , frequency-modulated by the moving-target bipolar video signals. This signal is amplified and limited and is then applied to the f.m. discriminator. This is a conventional circuit which provides a voltage output varying in magnitude and polarity in accordance with the variations in frequency of the input f.m. signal. Thus, the output from the demodulator block is the 'cancelled' bipolar video signal, i.e. a video signal which contains only *moving-target* information.

The final block in this delay-line canceller is the *bipolar rectifier*. The cancelled bipolar video from the f.m. discriminator is amplified and then applied to a full-wave rectifier. This rectifier is conventional and converts the bipolar video input to a *unipolar* form for intensity-modulation of the p.p.i. The cancelled unipolar video signal, containing only moving-target information, is then applied through a cathode- or emitter-follower output stage to the display unit.

Generation of PRF

We have seen that for perfect cancellation of fixed-object echoes in the delay-line canceller the delay time of the delay line and the pulse spacing must be *exactly equal*. We can usually achieve a fairly stable p.r.f., but the delay time in a delay line is more difficult to stabilize. The main reason for this is that the delay-line medium changes with variations in temperature.

There are two ways in which we can match the pulse spacing and the delay time:

- We can use a stable oscillator to generate the p.r.f. and make the delay line of variable length so that we can adjust the delay time to match the pulse spacing.
- The delay line can be included as part of the p.r.f. generator system. Any variation in the delay time will then automatically affect the p.r.f. of the system to keep the delay time and the pulse spacing exactly equal.

The second of these two methods is more common. The basic block schematic diagram of a typical system is illustrated in Fig 14.

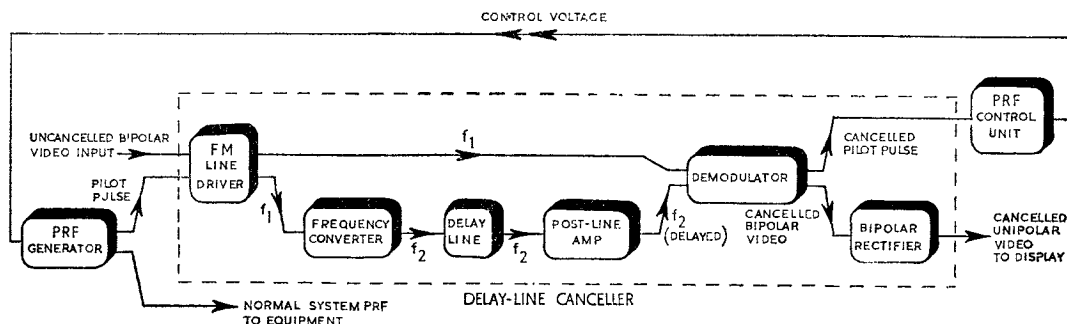


FIG 14. USE OF DELAY-LINE CANCELLER IN GENERATION OF PRF

The p.r.f. generator applies a 'pilot pulse' as an input to the delay-line canceller. If the

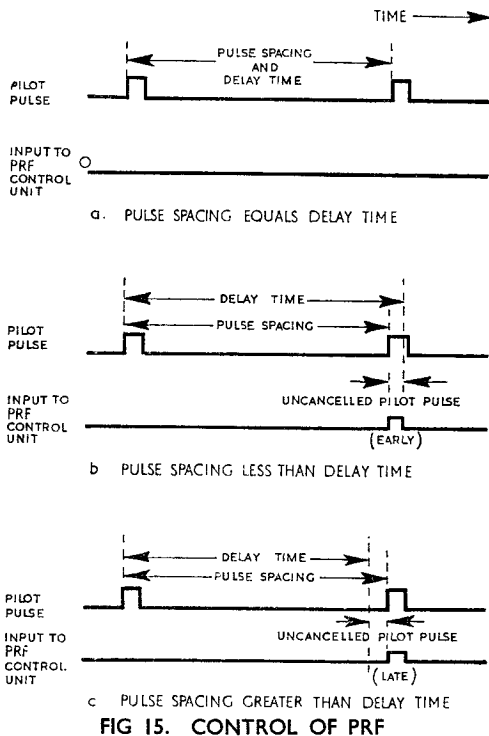


FIG 15. CONTROL OF PRF

If the delay time of the delay line and the pulse spacing are exactly equal, then the pilot pulse is completely cancelled at the output of the delay-line canceller. There is then no input to the p.r.f. control unit, and the p.r.f. generator is unaffected (Fig 15a). If, however, there is a discrepancy between the delay time of the delay line and the pulse spacing, there will be an uncanceled residue of the pilot pulse at the output of the delay-line canceller. This uncanceled pilot pulse is applied as the input to the p.r.f. control unit which then supplies a control voltage to the p.r.f. generator. This voltage is used to adjust the frequency of the p.r.f. generator in such a way as to bring the pulse spacing back into step with the delay time.

If the delay time of the delay line *increases*, the pulse spacing becomes effectively too small and an uncanceled pilot pulse appears at the output of the delay-line canceller *earlier* than it would normally do so (Fig 15b). If the delay time *decreases*, the pulse spacing becomes too great and the uncanceled pilot pulse appears *later* than it would normally do so (Fig 15c). By using a system of early and late gates we can control the voltage applied to the p.r.f. generator and thus automatically adjust its frequency as required.

Response of Delay-line Canceller

The delay-line canceller provides an output only for echoes which contain a Doppler shift component, i.e. for *moving-target* echoes. For fixed-object echoes (clutter) there is no Doppler shift and the output from the delay-line canceller is zero. Thus the delay-line canceller is, in effect, a *filter* which rejects the d.c. component of clutter.

However, there is another factor which determines the response of the delay-line canceller. If the target is moving with a relative velocity such that in an interpulse period the target moves a distance equal to a half-wavelength at the radiated frequency, then the pulse going out from the transmitter and returning will differ in phase from the previous pulse by *one complete wavelength*. Thus echoes from such targets will appear *not to have changed their phase* between successive firings of the transmitter and will be indistinguishable from fixed-object echoes. The response of the delay-line canceller will be zero, not only for the d.c. component of clutter (fixed objects), but also for those moving targets whose Doppler frequency happens to be the *same* as the p.r.f. or *harmonics* of the p.r.f. The relative frequency response of the delay-line is shown in Fig 16.

This response is a theoretical or *ideal* graph which assumes zero output at zero Doppler and at Doppler shifts equal to the p.r.f. or its harmonics. This is not so in practice. Clutter echoes are not always stationary; a tree blowing in the wind, although 'stationary', will produce an output from the delay-line canceller. Instabilities in the trans-

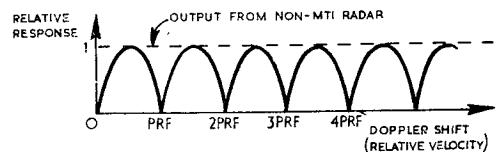


FIG 16. THEORETICAL RESPONSE OF DELAY-LINE CANCELLER

mitting or receiving equipment may produce an output for fixed objects. But the main cause of an output at the theoretical zero points arises from the *scanning action* of the aerial. As the aerial beam sweeps past an object the amplitude of the echo received from that object changes from pulse to pulse because of the shape of the beam. This gives rise to an apparent movement of a fixed object. The overall result is that the delay-line canceller produces an unwanted clutter output at the theoretical zero points (Fig 17).

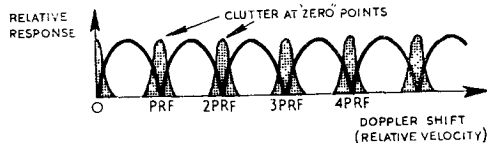


FIG 17. PRACTICAL RESPONSE OF SINGLE DELAY-LINE CANCELLER

To reject the clutter in the vicinity of d.c. or at Doppler frequencies corresponding to the p.r.f. and its harmonics, a *double cancellation* system may be used.

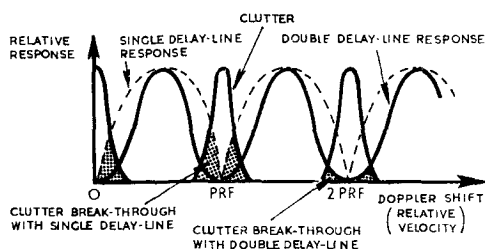


FIG 18. RESPONSE OF DOUBLE DELAY-LINE CANCELLER

The response of the double delay-line canceller is shown in Fig 18. The shape of the response curve is now modified so that more of the clutter in the vicinity of the theoretical zero points is rejected.

Ideally, any moving target should be visible over a fixed object, but because a fixed object leaves some uncanceled residue at the output of the canceller, this condition is not fully realized. The *sub-clutter visibility* is a measure of how closely the m.t.i. radar approaches the ideal. A sub-clutter visibility of 30db indicates that a moving target can be detected in the presence of clutter, even though the

clutter echo power is 1000 times the target echo power. Modern m.t.i. radars have values of sub-clutter visibility of the order of 20 to 40 db.

The *cancellation ratio* is a measurement of the effectiveness of the delay-line canceller in rejecting fixed-object echoes. It is the ratio of a fixed-object signal voltage *after* m.t.i. cancellation to the voltage of the same echo *without* cancellation. For a single delay-line canceller this ratio is of the order of 15 to 20 db. For a double delay-line canceller it is about 30 db.

Double Delay-line Canceller

A double delay-line canceller consists essentially of two single delay-line cancellers in cascade. The delay in each case must equal the pulse spacing of the radar and very often, to avoid inaccuracies, the *same delay line* is used for both cancellers. A basic block diagram is illustrated in Fig 19. There are two identical loops, the action of each being as described earlier

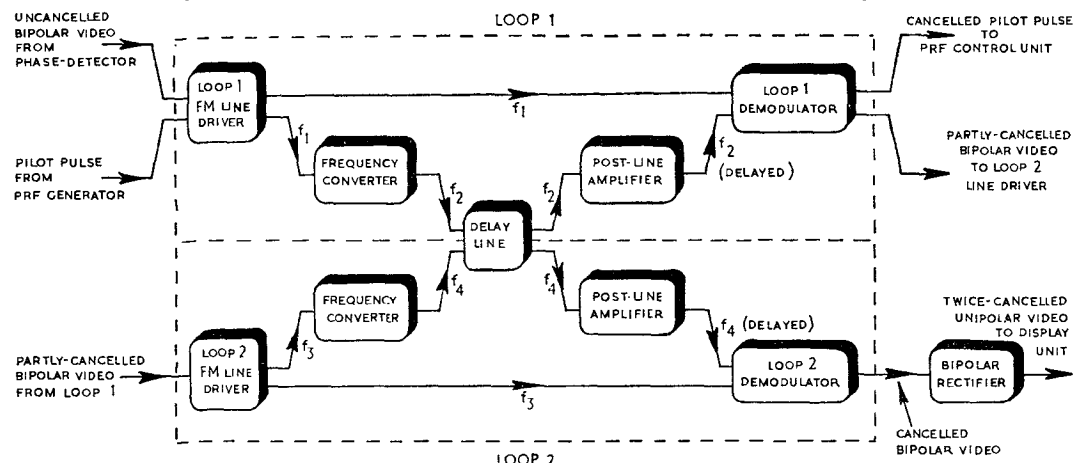


FIG 19. SIMPLIFIED BLOCK DIAGRAM OF DOUBLE DELAY-LINE CANCELLER

in connection with Fig 13. The fixed-object echoes are partly cancelled in loop 1 and the bipolar cancelled video output of loop 1 is fed as the modulating signal to loop 2 line driver. Further cancellation takes place in loop 2. The bipolar cancelled video output of loop 2 is then applied to the bipolar rectifier which provides the *twice-cancelled unipolar video output* for the display.

Blind Speeds

If we examine the frequency response curve of a delay-line canceller (see Fig 16) we shall see that for a given p.r.f. there are certain Doppler frequencies which give zero output. The relative velocities of a target corresponding to these Doppler shifts are known as *blind speeds*. At a blind speed the output of the delay-line canceller is the same as for a fixed object. Thus at certain target relative velocities we have *no indication of a moving target* and that radar is 'blind' at these target speeds.

From what has been said previously (see p 509) it should be clear that the blind speeds are related to the wavelength λ of the radar and also to the pulse repetition frequency f_r and harmonics of f_r . If λ is measured in centimetres, f_r in cycles per second and the relative velocity in knots, then the blind speeds v_n are given by:

$$v_n = \frac{n \lambda f_r}{102}, \text{ where } n = 1, 2, 3 \text{---}$$

Thus for a 10cm S-band radar operating with a p.r.f. of 1000 p.p.s., the first blind speed is about 98 knots (i.e. approximately one-tenth of the radar p.r.f.).

If the first blind speed is to be greater than the maximum relative velocity expected from the target then the wavelength λ must be large (low frequencies), or the p.r.f. f_r must be high, or both. However, as we have seen, the wavelength and the p.r.f. are usually decided by factors *other* than the blind speed. Therefore other means of increasing the first blind speed must be used.

We know that the blind speeds of two radars operating at the same p.r.f. will be different if their *frequencies* are different. It is therefore unlikely that both radars will be blind to the same moving target at the same time. *Duplication* of radars is a system which is sometimes used in m.t.i. radars. The duplication of the system also gives increased reliability.

Similar remarks apply to two radars operating at the same frequency but with *different values of p.r.f.* In this case, however, the same results could be obtained with a *single* radar using a *staggered* p.r.f. A radar with a staggered p.r.f. uses two or more different values of p.r.f., the radar being switched from one value to another at regular intervals of time—very often from pulse to pulse.

The effect of using a staggered p.r.f. is shown in Fig 20. Here we are using p.r.f.s of 622, 700 and 800 p.p.s., the corresponding pulse spacings being such that a stagger ratio of 9 : 8 : 7 results. The separate frequency response curve for each p.r.f. is shown and the bottom curve represents the combined response of the delay-line canceller. For a wavelength of 10cm, the first blind speed that is *common* to the three p.r.f.s in this example is 560 knots. This is greater than the maximum expected relative velocities of air-traffic-control targets. This figure can be raised even higher if we use *two radars* operating on slightly different frequencies with a common staggered p.r.f.

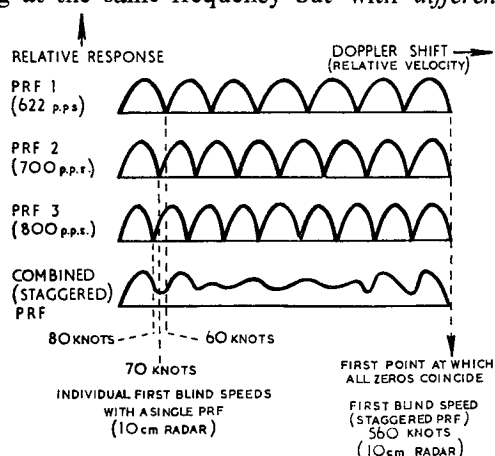


FIG 20. RESPONSE OF DELAY-LINE CANCELLER WITH STAGGERED PRF

Production of Staggered PRF

Fig 21 illustrates the basic idea of staggered p.r.f. production. The p.r.f. generator produces uniform trigger pulses with a pulse spacing of T_0 microseconds. These pulses are switched by

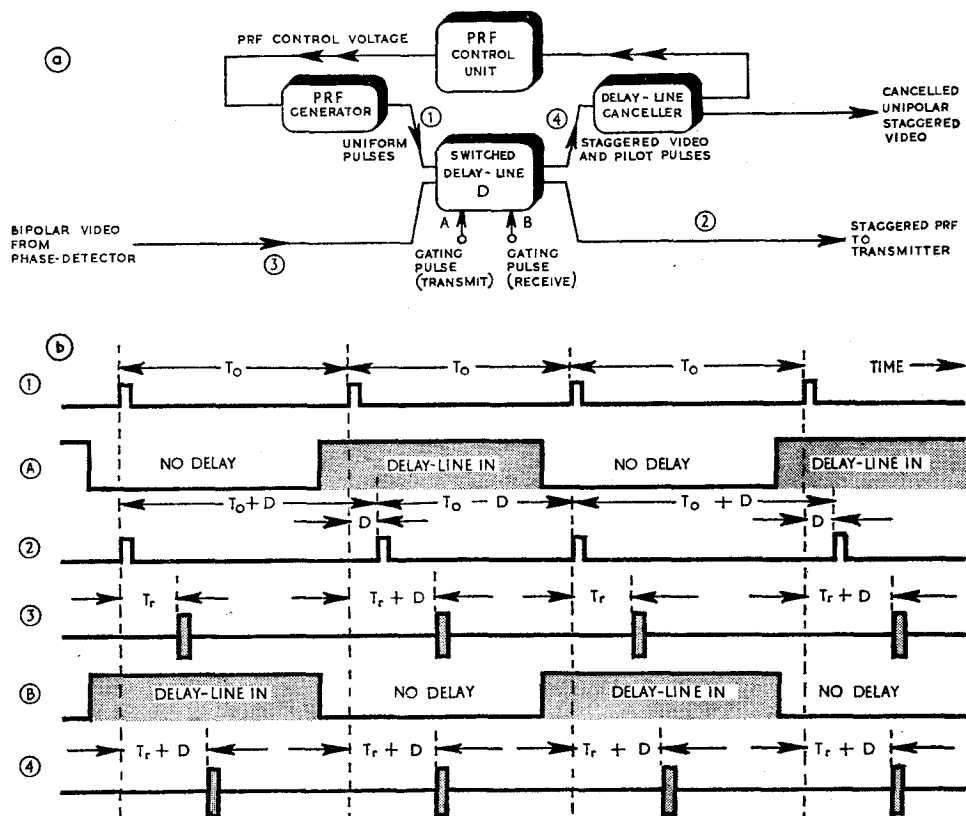


FIG 21. BASIC PRINCIPLE OF STAGGERED PRF PRODUCTION

a gating waveform in such a way that an undelayed path and a delayed path of time delay D microseconds are introduced alternately between the p.r.f. generator and the transmitter modulator. Thus the period between transmitted pulses is alternately $T_0 + D$ and $T_0 - D$ as shown in Fig 21b.

The same switched delay line is introduced into the receiver circuit. The target echo returns after a time T_r (depending on the range). The receiver gating waveform is such that during those periods when the transmitter pulse is delayed the received echo is undelayed. Conversely, when the transmitter pulse is undelayed, the received echo is delayed. This may be seen by comparing waveforms 3 and 4 which show the waveforms of the received signal at the input and output respectively of the switched delay line D . Hence the target echoes at the input to the delay-line canceller appear from pulse to pulse *at the same time* with respect to the uniform trigger pulses.

In a practical system a *three-pulse* staggered p.r.f. is often used. The basic outline of such a system is illustrated in Fig 22.

The trigger pulses from the trigger generator, occurring at uniform intervals of T_0 microseconds are applied to the f.m. line driver along with the uncanceled bipolar video signal from the receiver phase detector. The output from the f.m. line driver is a signal, frequency-modulated by the video and by the uniform trigger pulses (waveform 1). This output is divided

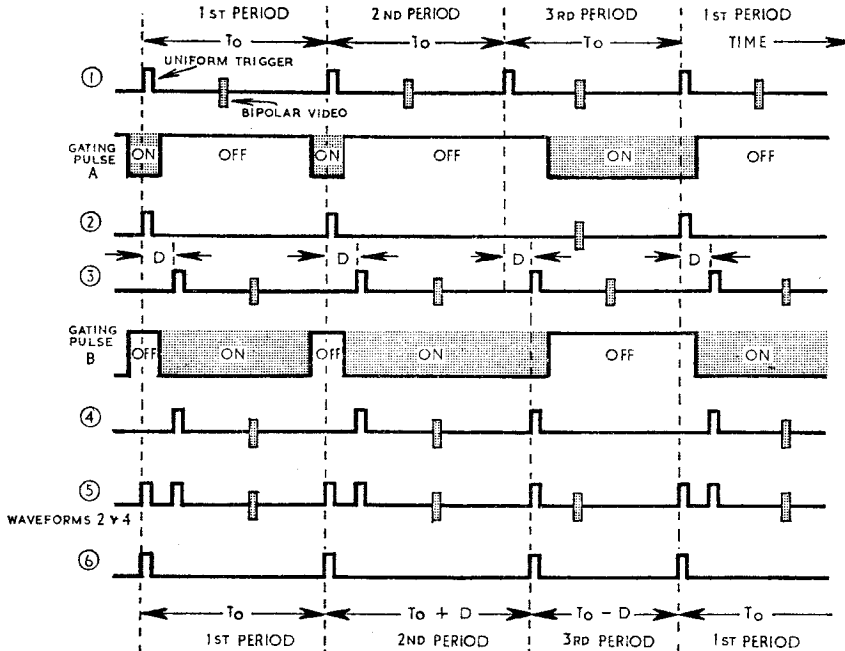
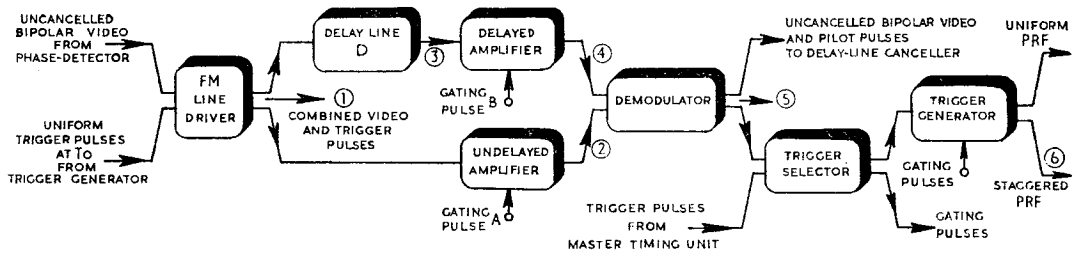


FIG 22. PRODUCTION OF THREE-PULSE STAGGERED PRF

between a delayed path and an undelayed path. The undelayed amplifier is gated by gating waveform A (supplied by the trigger selector) and this waveform is such that the output from the undelayed channel is represented by waveform 2. In the other channel the f.m. signal is delayed by D microseconds as shown by waveform 3. The delayed amplifier is also gated, this time by gating waveform B, so that the output from the delayed channel is as shown in waveform 4. The two inputs to the demodulator produce a combined demodulated output as shown by waveform 5. This output is applied to the delay-line canceller in the usual way so that cancellation of fixed-object echoes may be obtained. It is also applied to the trigger selector which removes the video and any unwanted triggers. The trigger selector provides the gating waveforms A and B and also fires the trigger generator. The latter provides output pulses at a waveform p.r.f. of pulse spacing T_0 microseconds and also the staggered p.r.f. with pulse spacing T_0 , $T_0 + D$, and $T_0 - D$ as shown by waveform 6.

We noted earlier in connection with Fig 21 that the video must be re-aligned with the staggered p.r.f. so that the target echoes at the input to the delay-line canceller appear from pulse to pulse at the same time with respect to the uniform trigger pulse. This is achieved in Fig 22 by the gating waveforms in conjunction with the delayed and undelayed paths.

Selective MTI

Although m.t.i. deals with the problem of permanent echoes by enabling the radar to distinguish between moving and fixed targets, this technique can only be applied where the aircraft has a radial motion component in relation to the ground radar. If the aircraft is flying *across* a radar sight-line it will appear to the radar as a fixed object (see p447). This means that an aircraft flying a certain course can disappear completely from the p.p.i.

The principle of selective m.t.i. (s.m.t.i.) is to limit the operation of m.t.i. to those areas where permanent echoes actually exist. SMTI provides a means of automatically switching between 'raw' radar video and cancelled video. This is accomplished using a flying spot scanner to view a photographic video map, on which the areas of permanent echoes are defined. Using the voltages thus produced to operate a switching circuit, raw or cancelled video can be automatically presented at any point on the p.p.i. as appropriate.

Summary

In this chapter we have considered the elementary principles of moving-target indication radar systems. Other systems are possible and have been used in practice. These systems include the use of banks of Doppler filters and range gates similar to that described for c.w. Doppler radars (see p449). A more modern development uses a computer to convert the video signals into digital form for subsequent subtraction and cancellation of fixed-object echoes.

However, enough has been said to show the possibilities and problems of m.t.i. radars. But, of course, for those employed on the maintenance of such equipments much more detailed information will be required. This will be found in the appropriate Air Publication for the equipment.

SECTION 8

RADAR DATA HANDLING

Chapter		Page
1	Outline of a Radar Defence and Control System	515
2	Microwave Radar Data Links	523
3	Analogue-to-digital Conversion	527
4	Radar Data Displays	537

OUTLINE OF A RADAR DEFENCE AND CONTROL SYSTEM

Introduction

A modern radar defence and control system has two main functions. Firstly, it collects and displays primary radar information, mainly on hostile targets, to provide quick and accurate defence information. Secondly, the primary radar information, plus information obtained from secondary radars, is used to control friendly aircraft within a given control space.

A defence and control system consists of *many* ground radar installations providing information on such things as range, bearing, height, and velocity of aircraft and missiles within its control space. Information from each separate radar site is passed to a *common control centre* where it is processed and displayed in a form which can be easily interpreted by the controllers. Rapid decisions can then be made to avoid collisions between friendly aircraft or to despatch fighters or missiles to intercept enemy targets.

The basic requirements of such a control and defence system are illustrated in Fig 1. Information from the radar sites is transmitted in suitable form along radar links to a link terminal

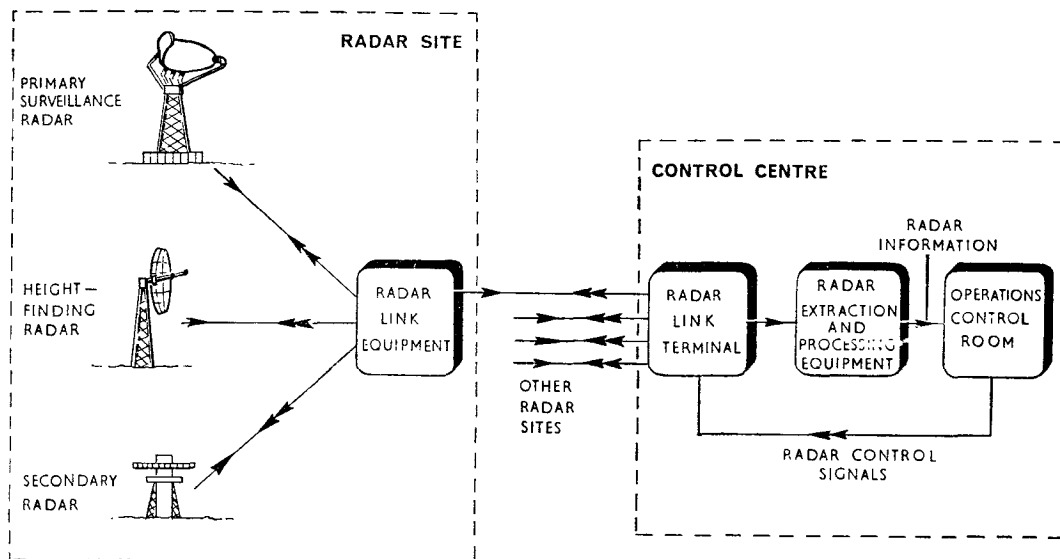


FIG 1. ELEMENTS OF A MODERN CONTROL AND DEFENCE SYSTEM

at the control centre. After extracting and processing the information, it is fed to the operations control room where it is displayed in various forms for use by the controllers.

Most of the information supplied by a radar is in *analogue* form, i.e. range may be represented by a voltage, and bearing and height by a shaft angle. Analogue information cannot easily be transmitted over long distances without loss of accuracy. Thus it is usually necessary to change the analogue information into *digital* form at the radar before transmission to the control centre. In this section we shall consider how the analogue data obtained from the radar are changed into digital data, the methods of transmitting these data, and the ways in which the data are processed and displayed to the controllers.

Radar Data Links

In a complex radar surveillance and control system it is not often possible to site the radar near the operations centre. The radar must be sited where the best possible echoes can be obtained, free from ground clutter, blind areas and interference sources. The control centre would be placed where it can be linked conveniently with several radar sites and airfields.

When several radars are situated within a few miles of each other they can cause mutual interference. This can be reduced by synchronizing the various p.r.f.s and also the rotation of the aerials. The synchronizing signals can be originated at the control centre. Further, it is often desirable for the radar operator at the control centre to be able to control circuits such as variable gain, anti-clutter and anti-jamming circuits in long-range search radars. The controller may also want additional information from the primary and secondary radars and he must be provided with facilities so that he can *demand* this information.

Thus the link between radar site and control centre is required to carry radar information *from the site* to the control centre and also *control* instructions *from the control centre* to the radar (Fig 2).

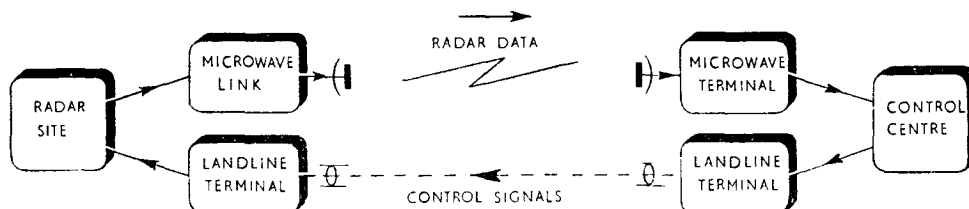


FIG 2. COMMUNICATION LINKS IN RADAR DATA SYSTEM

The more information conveyed by the radar link, the wider must be the link *bandwidth*. All the information obtained by the radar at the radar site, including clutter and unwanted echoes, is called *raw radar*. Some of this raw radar information is not required for display at the control centre and may therefore be omitted, thus reducing the bandwidth of the link channel. The minimum information required at the control centre includes the radar trigger and timing pulses, the radar video pulses, and the aerial bearing and elevation angles.

The position of a target on a local raw radar display is given in polar co-ordinates $r \angle \theta$, where r is the range and θ the bearing from the radar site. This can be transposed into x and y cartesian co-ordinates and *corrected* so that the position of the target *relative to the control centre* is obtained (Fig 3). This simplifies the overall picture of aircraft movements to the controller.

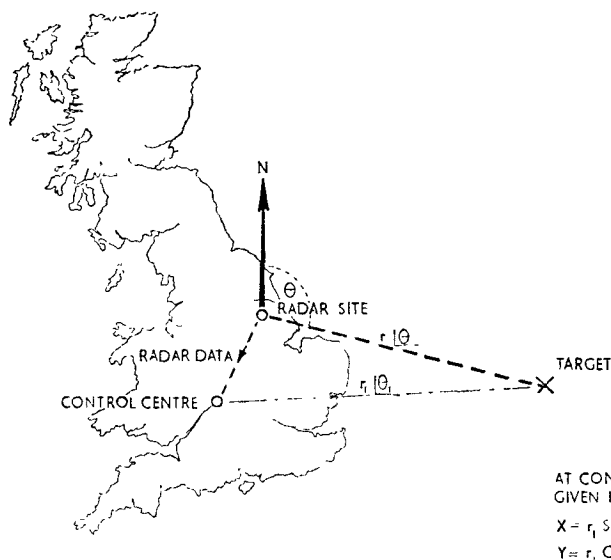


FIG 3. CONVERSION OF RANGE AND BEARING TO CORRECTED CARTESIAN CO-ORDINATES

The links between the radar sites and the control centre can be either cable or radio links. Very often a mixture of both is employed. Wideband signals, such as video and triggering pulses, are usually sent by *radio* link. The narrow-band demand and control signals are usually sent over *landline*. Coaxial cables are normally used for landlines and they may be ducted, buried, or supported on poles. The cables must be screened against interference. Over the longer lengths of cable run, amplifiers and repeating equipment may be necessary.

To accommodate the wide bandwidth necessary for the transmission of the video and triggering pulses, the radio link used operates on frequencies between 1000 and 8000 Mc/s. The system is then called a *microwave link*. Distances of 50 miles can be covered and, by using repeater stations, distances of 200 miles and more are possible.

Digital Data Transmission

The elevation and bearing information obtained from a radar is in *analogue* form, i.e. it is represented by the angular position of a shaft relative to a given datum. In this form the information can be transmitted over hundreds of yards by means of synchros. If we wish to transmit the information over many miles, and retain a high degree of accuracy, it must be converted from analogue to *digital* form.

Similarly, a d.c. voltage representing the range of a target, if sent directly along the transmission path, would be attenuated and inaccurate information would be received. By converting the d.c. voltage into digital form representing its polarity and magnitude, and transmitting the information in this form, greatly improved accuracy is obtained (Fig 4).

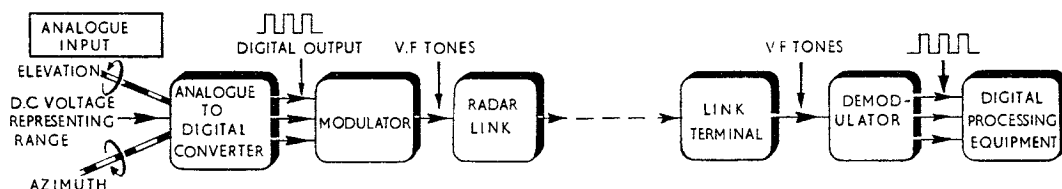


FIG 4. CONVERSION AND TRANSMISSION OF RADAR DATA

To send digital information over long distances by landline or radio, some type of modulation is required (Fig 5). In amplitude modulation, the two levels of the digital pulse are represented by different amplitudes of sine wave (Fig 5b). In frequency modulation, the sine wave frequency is varied (Fig 5c), and for phase modulation the phase of the sine wave is changed (Fig 5d).

The information supplied by a radar site may include *several* video signals (from different radars on the site), several triggering and timing pulses, and separate bits of digital information on range, bearing and height of targets. The *secondary* radars on the site may provide additional information on identity of targets, their route, flight-plan, and so on. To transmit all this different information from the site to the control centre, *multi-channel carrier telegraphy* may be used (see p190 of AP3302 Part 2).

In this system each separate bit of information is used to frequency-, amplitude- or phase-modulate a signal which is then *translated* in frequency by the injection of a carrier signal to provide one 'sub-channel'. The different

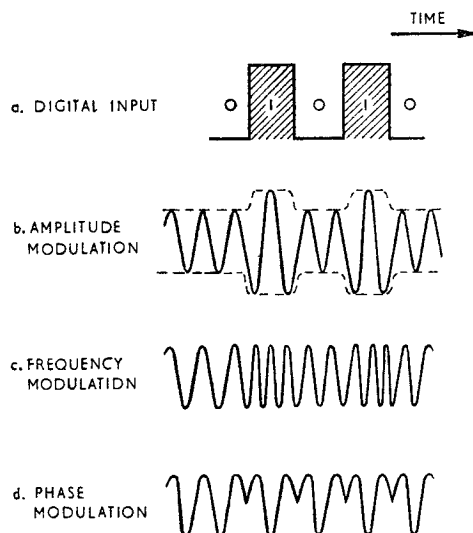


FIG 5. METHODS OF MODULATION

TYPICAL RADAR
INFORMATION IN
DIGITAL DATA FORM

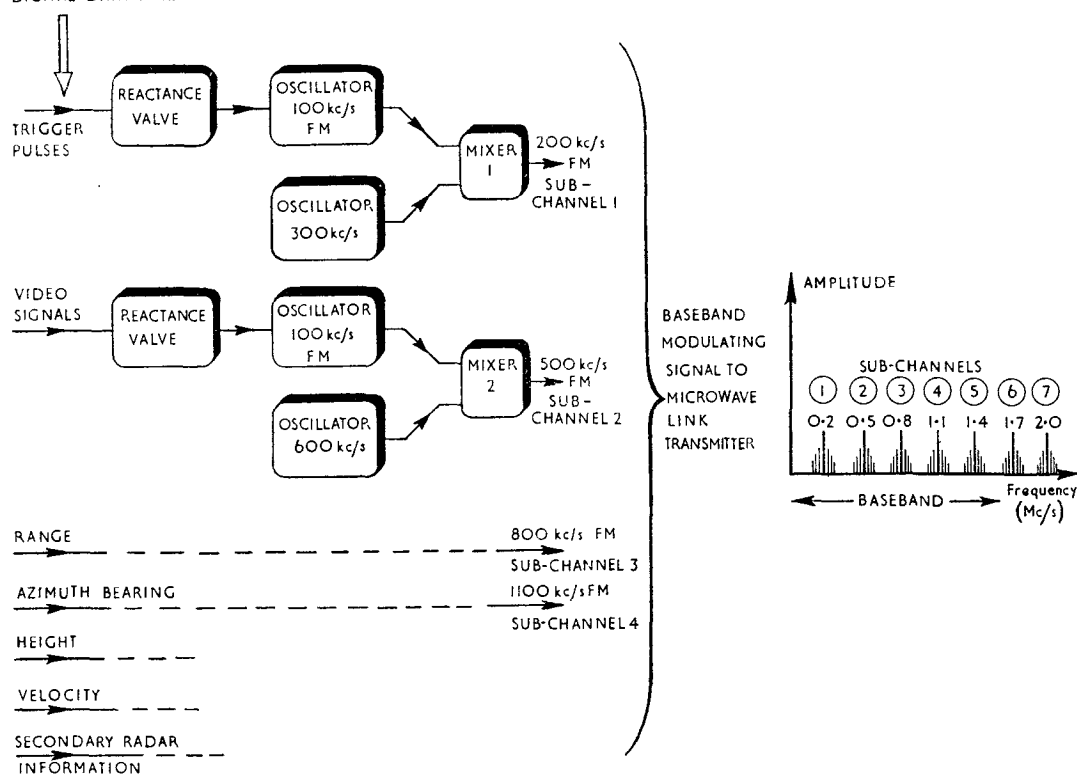


FIG 6. USE OF MULTI-CHANNEL CARRIER TELEGRAPHY FOR TRANSMISSION OF RADAR DATA

bits of information are treated in the same way but each has a different injected carrier frequency so that different sub-channels, spread over a range of frequencies, are produced (Fig 6). This method is similar to the *baseband* system discussed in Part 2 of AP3302. The baseband, which includes all the separate modulated sub-channels, is used to modulate the link transmitter. We can see that the baseband may be spread over several Mc/s and to accommodate such a wide bandwidth a microwave link is *essential*. At the receiving end (the control centre) the reverse process takes place to extract the various bits of radar information.

If we assume that the microwave link has a nominal carrier frequency of 2000 Mc/s, the radiated signal from the site may cover the band 2000 to 2008 Mc/s. However, we have several radar sites all feeding into a common control centre. Thus, to avoid mutual interference, the baseband modulated signal from each radar site is centred on a slightly different microwave carrier.

Data Processing

By converting the radar information into digital form we improve the accuracy of transmission. However, digital data transmission has another advantage: the received information can be quickly and accurately processed in a digital computer. This relieves the controller of cumbersome calculations and enables him to make quick decisions based on accurate and up-to-date information.

For example, in a ballistic missile early warning system, the target range, bearing, height and relative velocity can be digitized and fed as binary numbers into the store of a digital computer.

By tracking the target for a few seconds a chain of such numbers can be built up and the computer can calculate the trajectory of a missile from these numbers. Further calculation by the computer can then give the impact point and impact time of the missile.

With surveillance and height-finding radars, digitized information from a large number of targets can be fed into a digital computer. The computer then calculates predicted positions, intercept tracks and so on, thus presenting the controller with accurate information with which to control defending aircraft and missiles.

Thus, one of the functions of a radar *processing* equipment is to calculate the track of each target, store it, and keep it up to date by using fresh information provided by the primary radar (Fig 7). This information can be read out for *marking* the p.p.i. radar displays or for use in *tabular* displays.

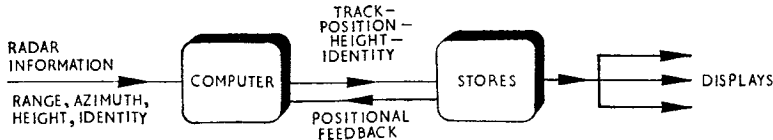


FIG 7. PROCESSING OF RADAR DATA AT CONTROL CENTRE

Secondary Surveillance Radar

The function of primary radar is to supply information covering targets which *do not co-operate* with the ground radar. This information is normally limited to range, bearing, height and velocity. From this, further data such as track can be computed. To obtain additional or more accurate information *secondary radar* is required.

Secondary radars form an important part of a defence and control system. Height-finding primary radars are not usually sufficiently accurate to indicate the 1000-feet separation between aircraft required in air traffic control. By feeding height information directly from the aircraft barometric altimeter into a secondary radar transponder fitted in the aircraft, the transponder, when interrogated by a ground radar pulse, can supply accurate and up-to-date height information in digital form. Other information such as aircraft identity can be extracted from the transponder without reference to the aircrew.

A fully automatic secondary radar system would thus form an automatic air-to-ground data link. Details of the aircraft type, proposed route, height, estimated time of arrival at reporting points and so forth are filed by the aircrew before departure. This flight plan information is digitized and fed into a computer at the departure control centre and is transmitted to computers at intermediate and destination control centres. The flight plan can be corrected and brought up to date during flight by additional information obtained from primary and secondary radars (Fig 8). Thus, a radar controller could 'strobe' a particular aircraft echo on

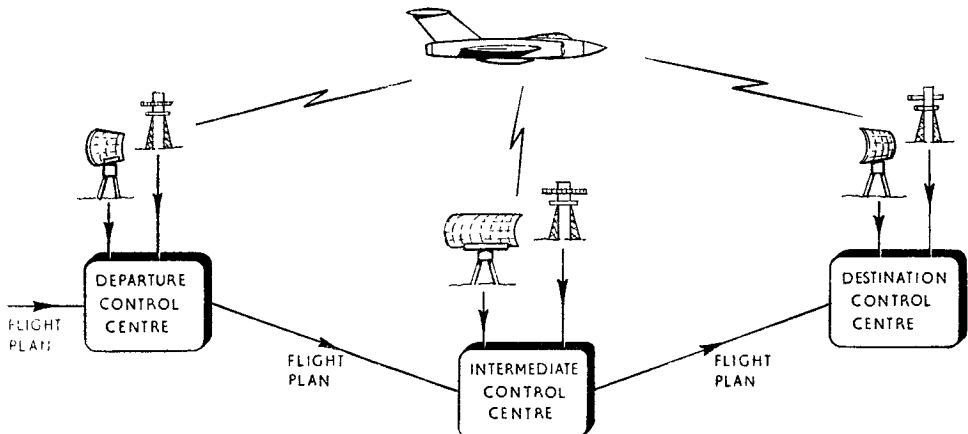


FIG 8. CONTINUOUS UP-DATING OF FLIGHT PLAN

his p.p.i. display and have the associated flight plan information immediately displayed. Alternatively, if a particular flight plan is selected, the aircraft to which it relates can be indicated on the p.p.i. display.

A further advantage of secondary radar over primary radar, when considering an automatic data processing control system, is that there is no ground or weather clutter present with the returned pulse. Therefore the coded information from the transponder can be decoded and fed directly into a computer or through a *pattern generator* on to a display.

The interrogation signal from the ground radar consists of two or three pulses whose time separation defines the question to be answered. The system can have several of these 'modes' each corresponding to a definite question such as 'What is your height' or 'What is your identity'. The reply signal from the aircraft consists of a number of coded pulses, each accurately timed. This signal is passed through decoding equipment into a computer, and from the computer store to various types of display for presentation to the controller.

Displays

In a control centre which handles a large number of aircraft there would be many radar controllers, each responsible for the aircraft in a given air space. Each controller would have a number of displays conveniently mounted on a console. The p.p.i. display which shows all targets in the controller's sector is the *raw radar* display. The aircraft echoes on this display have not been 'processed', i.e. they have not passed through data processing equipment. Thus, the raw radar display contains some information of little interest to the controller and, at the same time, it lacks information such as track and identity of the targets with which he is concerned (Fig 9).

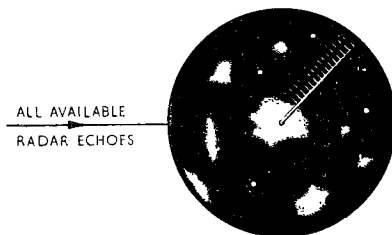


FIG 9. RAW RADAR DISPLAY

When radar data have been processed by a digital computer they can be held in a store. The information in the store can be scanned and data concerning position, track, height and identity of a target can be extracted. Thus a controller can demand information from the store and the required information about a target appears on the p.p.i. display of the controller initiating the demand. This type of display is called a *synthetic* display; it only shows information concerning those targets *requested* by the controller (Fig 10).

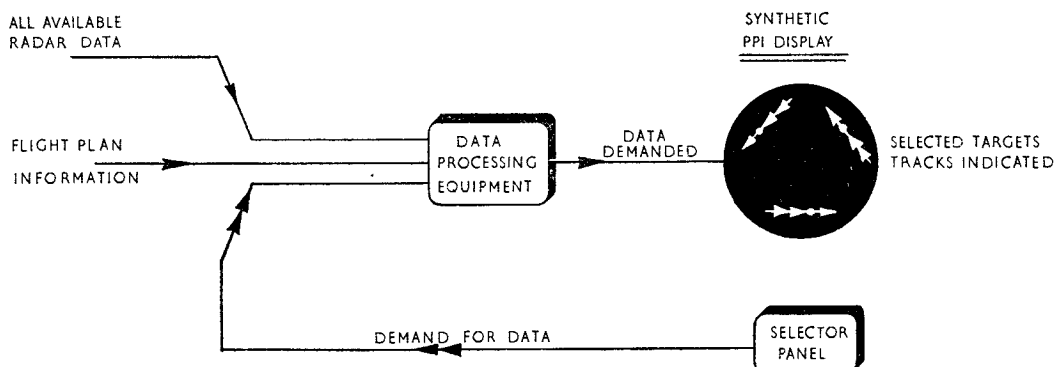


FIG 10. SYNTHETIC DISPLAY OF RADAR INFORMATION

The information held in the computer store can also be changed into letters and numbers and displayed on a c.r.t. screen in *tabular* form. Information on this tabular display is automatically up-dated as data concerning each target changes and only that information demanded by the controller is displayed (Fig 11).

To enable a controller to associate echoes on his p.p.i. screen with important geographical features such as coastlines, and reference points such as reporting beacons, a *video map* can be traced on the p.p.i. radar display (Fig 12). This is a map of the area of interest to the controller. The information shown on the map is mixed with the radar video input and applied to the control grid of the display tube.

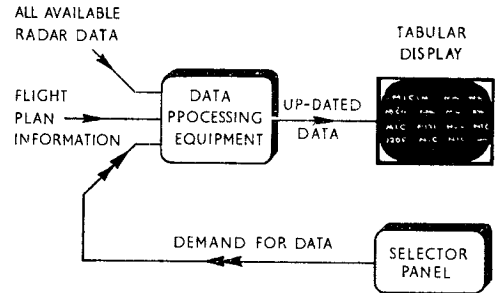


FIG 11. TABULAR DISPLAY OF RADAR INFORMATION

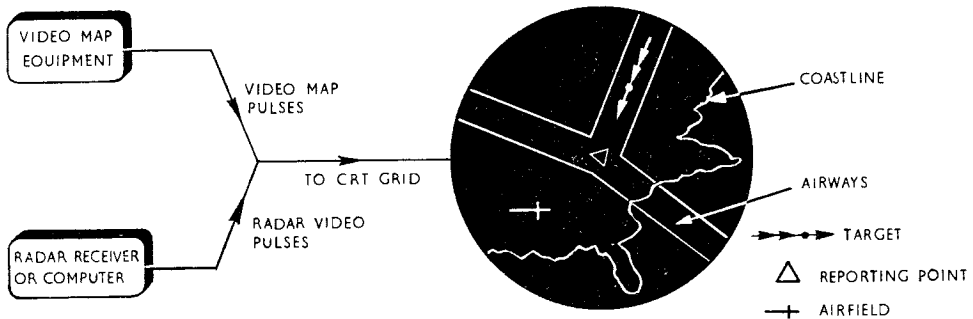


FIG 12. USE OF VIDEO MAP IN RADAR PROCESSING

In a control centre which is not fully automatic, detailed information about each aircraft can be displayed by means of *closed circuit television*. The information is written on a perspex board behind which is mounted a television camera. There are several such boards and cameras and the controller can select the one which contains information of interest to him.

A more recent development is known as *photo-display*. This uses a combination of coloured photographs and the processed radar information to provide a three-dimensional (3D) picture of the air space of interest to the controller.

MICROWAVE RADAR DATA LINKS

Introduction

In Chapter 1 we have seen the need for radio links between the radar sites and the control centre. The wideband video and triggering pulses and the digital data pulses are combined in a *baseband modulating system* and linked with the control centre by microwave links. Microwave radar links operate on selected frequencies in the 1000 to 8000 Mc/s band and because of these super-high frequencies a very wide bandwidth is available. Small aerials mounted on towers give very narrow beams resulting in line-of-sight reception between transmitter and receiver. Because of the narrow beams, low-power transmitters can be used; a transmitter output of 1W gives adequate reception at a distance of 50 miles. Where line-of-sight transmission between two points is not possible because of obstacles (e.g. mountains), *repeater stations* can be inserted between the radar site transmitter and the control centre receiver.

Basic Microwave Link

A block diagram illustrating the basic idea of a microwave radar data link is shown in Fig 1. Any number of repeaters may be used depending upon the distance being covered and the obstacles between the terminals. The repeaters receive the weak signal from the previous station, amplify it, and re-transmit it to the next station. To avoid interference between received and transmitted signals at the repeater, *slightly different frequencies* are used for reception and transmission.

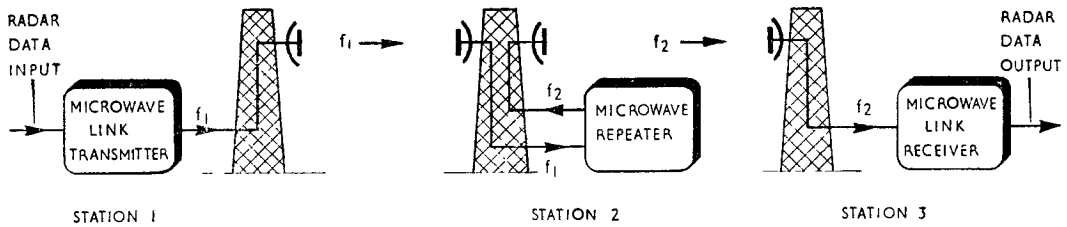


FIG 1. OUTLINE OF MICROWAVE RADAR DATA LINK

Transmitter

One form of microwave link transmitter is shown in Fig 2. The baseband input signal, which contains all the relevant data from the radar site, is amplified in the modulator stage. The amplified modulating signal is then used to amplitude-modulate the s.h.f. carrier from the oscillator. The oscillator operates at the output microwave frequency and could be a reflex klystron or a low-power travelling-wave tube. The amplitude-modulated signal is then amplified in a power travelling-wave tube and applied to the aerial.

A *frequency-modulated* microwave link transmitter is illustrated in Fig 3. The baseband radar signal is amplified in the modulator and then applied to a reactance valve to modulate the f.m. oscillator. The f.m. oscillator operates at a fairly low frequency (e.g. 60 Mc/s). To provide the required output microwave frequency

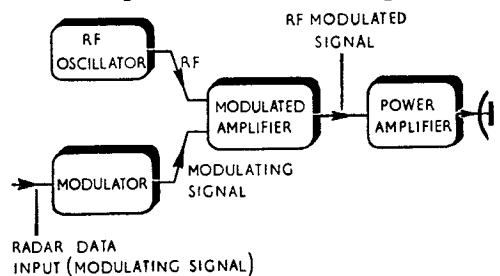


FIG 2. BASIC BLOCK DIAGRAM OF AMPLITUDE-MODULATED MICROWAVE LINK TRANSMITTER

the f.m. signal from the f.m. oscillator is mixed with the output of an s.h.f. oscillator (reflex klystron or backward-wave oscillator) in a travelling-wave tube frequency-changer. The output from this is the required f.m. microwave signal which is further amplified by a power travelling-wave tube before being applied to the aerial.

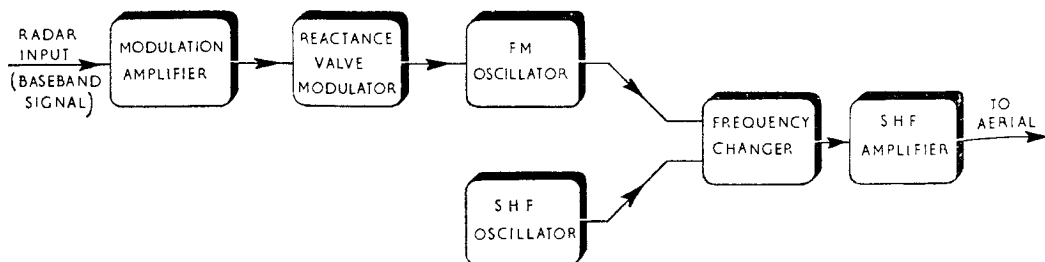
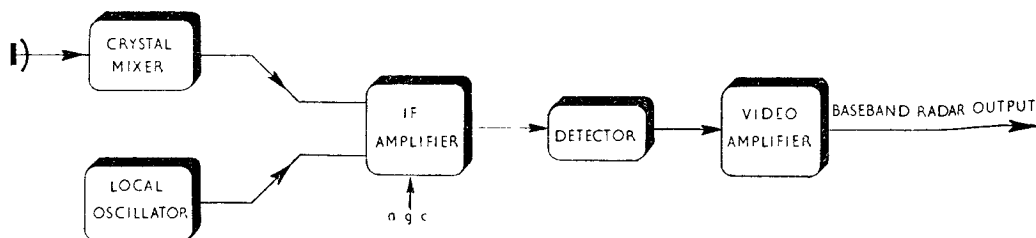


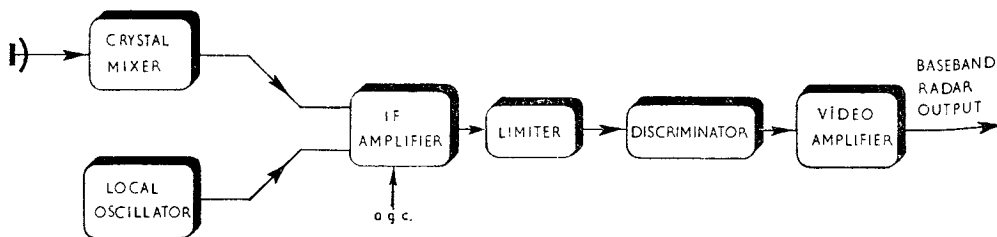
FIG 3. BLOCK DIAGRAM OF FREQUENCY-MODULATED LINK TRANSMITTER

Receiver

A microwave link receiver takes the conventional superhet form shown in Fig 4a for an amplitude-modulated signal and that shown in Fig 4b for a frequency-modulated signal. The crystal mixer has a low noise factor and the local oscillator is a reflex klystron, a coaxial-line oscillator, or a travelling-wave tube oscillator. AGC is provided to combat fading. To increase the signal-to-noise ratio a low-noise travelling-wave tube may precede the mixer stage and act as a microwave signal amplifier.



(a) AMPLITUDE MODULATION



(b) FREQUENCY MODULATION

FIG 4. MICROWAVE LINK RECEIVER, SIMPLIFIED BLOCK DIAGRAMS

Repeaters

The repeater stations are situated between the radar site and the control centre and receive the weak signal from the previous station, amplify it, and re-transmit it to the next station. It is worth noting that a repeater does not have to *demodulate* the signal during its passage through.

Usually, the incoming signal is converted to an intermediate frequency, for ease of amplification, and then, after adequate amplification, converted to the output frequency for onward transmission.

A block diagram of one form of repeater is shown in Fig 5. The input signal frequency f_1 (4000 Mc/s in Fig 5) differs from the output frequency f_2 (4040 Mc/s) so that the stability of the repeater is improved.

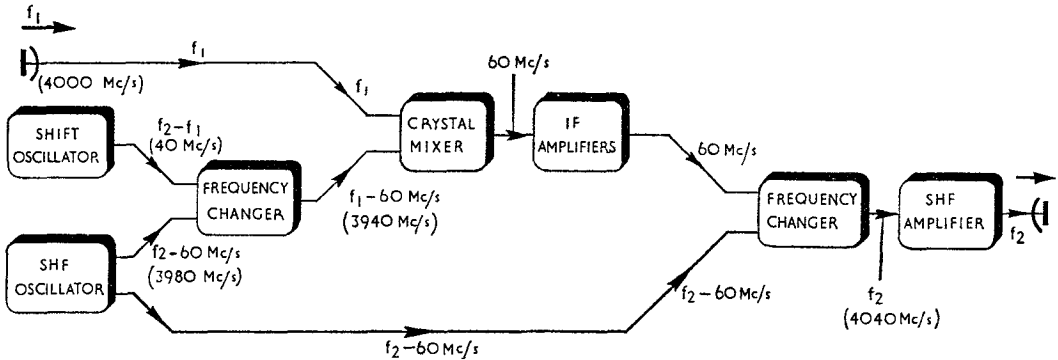


FIG 5. MICROWAVE LINK REPEATER, BASIC BLOCK DIAGRAM

The i.f. of 60 Mc/s is produced by mixing the input signal f_1 with a frequency 60 Mc/s below it. The latter is obtained by mixing the output from an s.h.f. oscillator operating at the output frequency f_2 minus 60 Mc/s with the output from a crystal-controlled shift oscillator working at the difference frequency between output and input ($f_2 - f_1$). The resulting i.f. signal from the crystal mixer is amplified and then converted to the final output frequency f_2 by mixing it with the s.h.f. oscillator output. Further amplification takes place at the microwave output frequency before the radar-modulated signal is re-transmitted.

The frequency-changers may be crystal diodes or travelling-wave tubes; the i.f. amplifier is a normal wideband amplifier with a bandwidth capable of handling the baseband signal; and the s.h.f. amplifier is a multi-cavity klystron or travelling-wave tube.

The development of low-noise travelling-wave tube (t.w.t.) amplifiers has enabled amplification to be carried out *at s.h.f.* Thus by using t.w.t.s it is possible to design a repeater in which the received signal does not need to be reduced to an i.f. before amplification occurs. Fig 6 shows the block diagram of an all-t.w.t. repeater. This is a much simpler arrangement than that of Fig 5 and contains fewer components. Filters are used to reduce the bandwidth accepted by the t.w.t. and improve the noise factor. A crystal-controlled oscillator and a t.w.t. frequency-changer are used to change the frequency of the received signal before re-transmitting it.

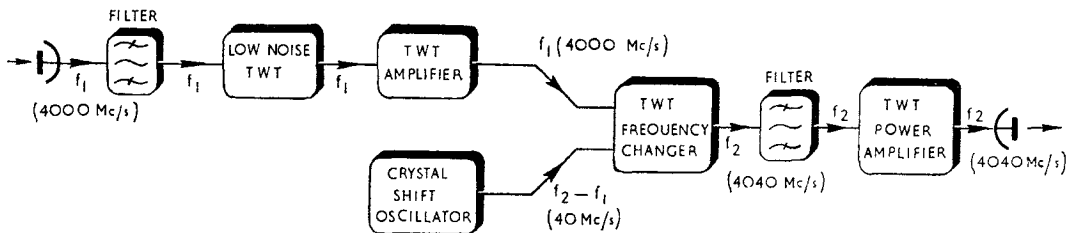


FIG 6. MICROWAVE LINK REPEATER USING TRAVELLING-WAVE TUBES, BLOCK DIAGRAM

Repeater stations are usually unattended. If the main power supply fails, diesel-electric auxiliary supply sets immediately start up and provide power until the main supply is again available.

Solid State Microwave Link

A microwave link, consisting of a transmitter and receiver in which only *solid state devices* are used, is shown in Fig 7. It operates in the 2 Gc/s band and has a transmitter power output of 2W. The transmitter (Fig 7a) consists of a transistor crystal-controlled oscillator-frequency multiplier, the output of which (at 125 Mc/s) is amplified by transistor amplifier stages to produce a power output of 25W at 125 Mc/s. This output is then applied to four varactor frequency-doubler stages, the final output being a 5W signal at 2 Gc/s. This output is passed to a transistor-varactor stage where it is frequency-modulated by the baseband radar data before being applied to the aerial.

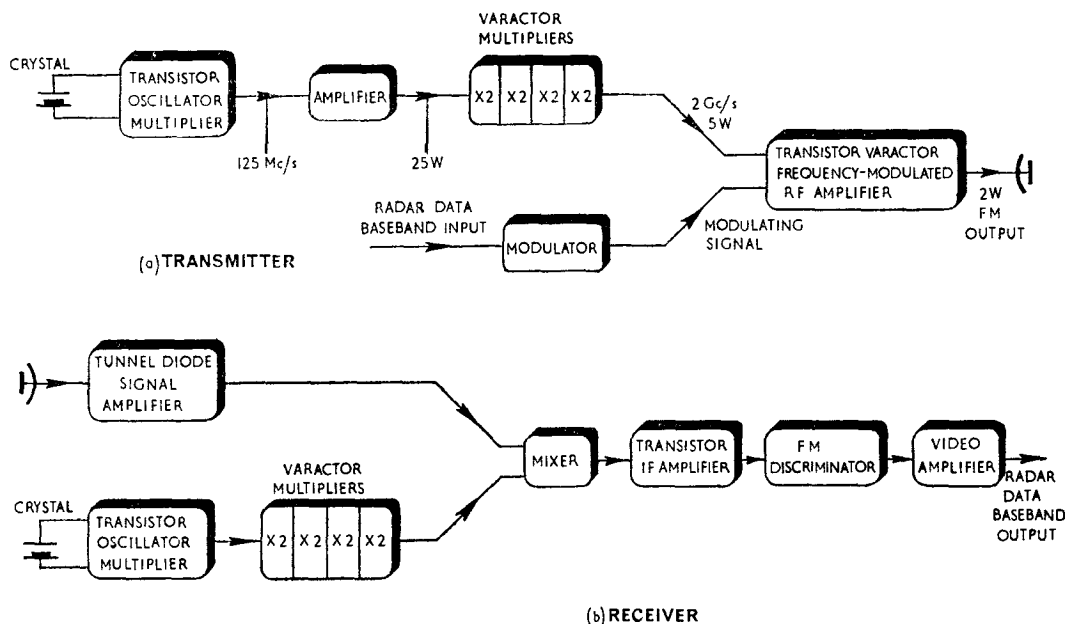


FIG 7. SOLID STATE MICROWAVE LINK EQUIPMENT, BLOCK DIAGRAM

The first stage of the receiver (Fig 7b) is a microwave tunnel diode which amplifies the incoming signal at 2 Gc/s. This stage has a gain of 14 db and a noise factor of 5 db. The local oscillator consists of a transistor crystal-controlled oscillator followed by four varactor frequency-doublers. It produces a 100 mW signal at a frequency which differs from that of the incoming signal by the i.f. After mixing in a crystal diode the i.f. signal is extracted and amplified in a wideband i.f. amplifier consisting of sixteen transistor stages, with a total gain of 75 db. The baseband video is extracted by an f.m. discriminator and applied to transistor video amplifier stages.

The equipment operates from a 24V d.c. supply and can be connected across a float-charged battery. Thus temporary failure of the mains supply does not affect operation of the link. Because of the solid state nature of the equipment, reliability is high and servicing is reduced to a minimum.

Propagation

The aerials used in a microwave radar data link are normally parabolic dishes fed by a dipole or by a waveguide horn. Thus the microwave energy is concentrated into a narrow pencil beam. In repeater stations the transmitter and receiver aerials are usually mounted back to back.

At s.h.f. any change in the refractive index of the atmosphere due to a change of air density or water vapour content can cause the radio beam to be bent and this produces fading at the receiver. These fades are most noticeable at dusk and dawn and are more severe where the propagation path is over water. The degree of fading depends upon the height of the aerial masts and upon the frequency of transmission, and these factors are chosen to reduce the effects of fading to a minimum.

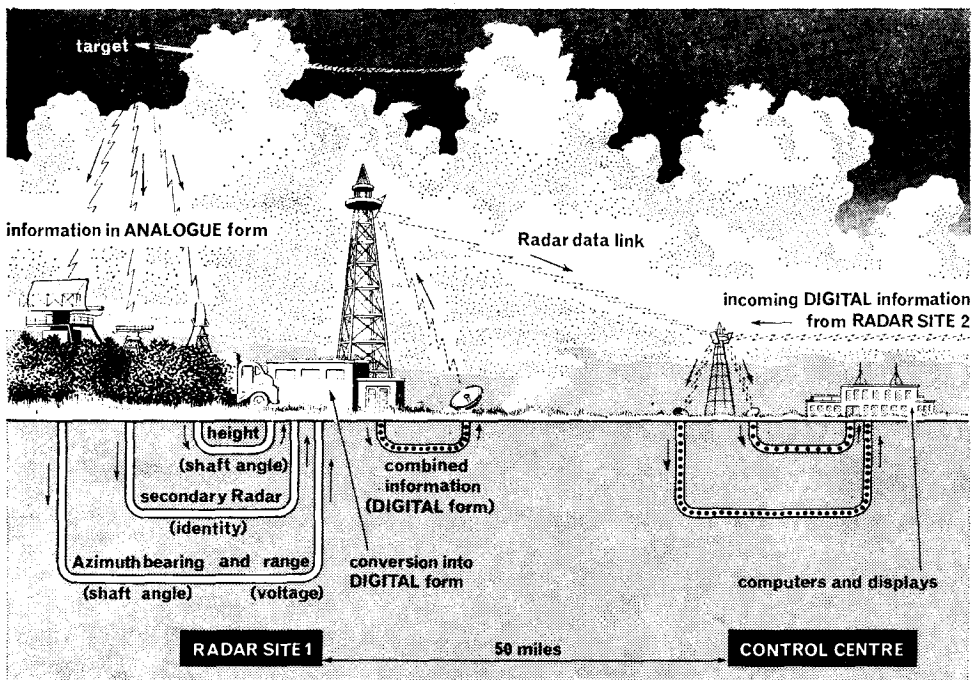
ANALOGUE-TO-DIGITAL CONVERSION

Introduction

We have seen earlier that the output of a radar is normally in analogue form. The target range may be represented by a voltage of a certain magnitude and polarity; the elevation or bearing of the target may be represented by the angular position of a shaft; and so forth. We cannot easily transmit analogue information over distances greater than a few hundred yards without loss of accuracy. Because of this it is usual to convert the analogue information into digital form before transmitting it over long distances. Thus a voltage of $+12V$ representing, say, a range of 53 miles could be converted into a *number* corresponding to this range; a shaft rotation of 20° from due North could be converted into a number representing this bearing. The number is normally transmitted in a form of *binary code*, where a pulse represents the digit '1' and the absence of a pulse the digit '0'. Details of the binary code and some of the circuits we shall discuss in this chapter are given in the chapter on Digital Computers in Part 1 of these notes.

After passing over the microwave link in digital form the number is received and converted back into its analogue form for presentation on a radar display; or, still in its digital form, it can be fed to a computer where, with other information, it can be used to calculate further information about the target.

In this chapter we shall consider methods of converting analogue quantities into digital form. There are normally two problems; the conversion of a shaft angle analogue into its digital equivalent; and the conversion of a voltage analogue.



Conversion of Angle Analogue into Digits

To convert a shaft angle into its digital equivalent, the 360 degrees through which the shaft can turn is divided into a number of sectors, where each sector is represented by a digital number. This can be done by mounting a number of differently-shaped cams on the shaft as shown in Fig 1. The drive shaft may be an extension of the aerial shaft. In this case the analogue input is the bearing angle of the aerial and we require to convert this to digital form.

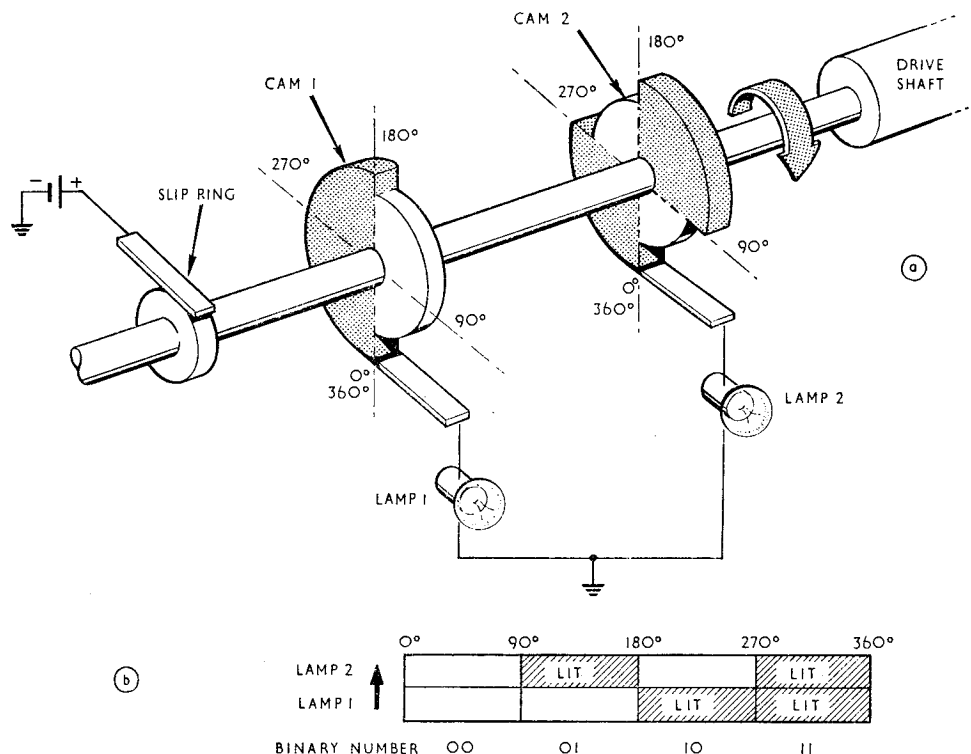


FIG 1. CONVERSION OF SHAFT ANGLE TO BINARY DIGITS, BASIC IDEA

In Fig 1 a voltage is applied to the shaft via a slip ring and the cams are cut in such a way that they each make or break a circuit to a lamp, depending upon the shaft angle. Cam 1 is cut so that the contact to lamp 1 is broken between a shaft angle of 0° and 180° (unshaded portion of cam). Between 180° and 360° (shaded portion) the contact is made and lamp 1 lights. Thus lamp 1 indicates in which of these two sectors the angular position of the shaft lies.

In a similar way we can see that cam 2 breaks the circuit to lamp 2 between 0° and 90°, makes it between 90° and 180°, breaks it between 180° and 270° and makes it between 270° and 360°. The combined information supplied by the two lamps indicates one of four sectors in which the shaft lies. As the drive shaft rotates, the combination changes every 90° and the lamps indicate the angular position of the shaft to within 90°.

If we assume that a lit lamp represents the digit 1 and an unlit lamp the digit 0, then the 0° to 90° quadrant is represented in binary form by 00, the 90° to 180° quadrant by 01, the 180° to 270° sector by 10, and the 270° to 360° sector by 11 (Fig 1b).

By adding more specially-shaped cams to the drive shaft the accuracy of shaft position can be improved. For example, with *four* cams the shaft position can be read to within $22\frac{1}{2}^\circ$. Fig 2 shows the pattern which would be produced by four cams cut in accordance with the binary code. As the shaft rotates the lamps go on and off according to the cam pattern, so indicating in binary code which of the sixteen sectors the shaft is in.

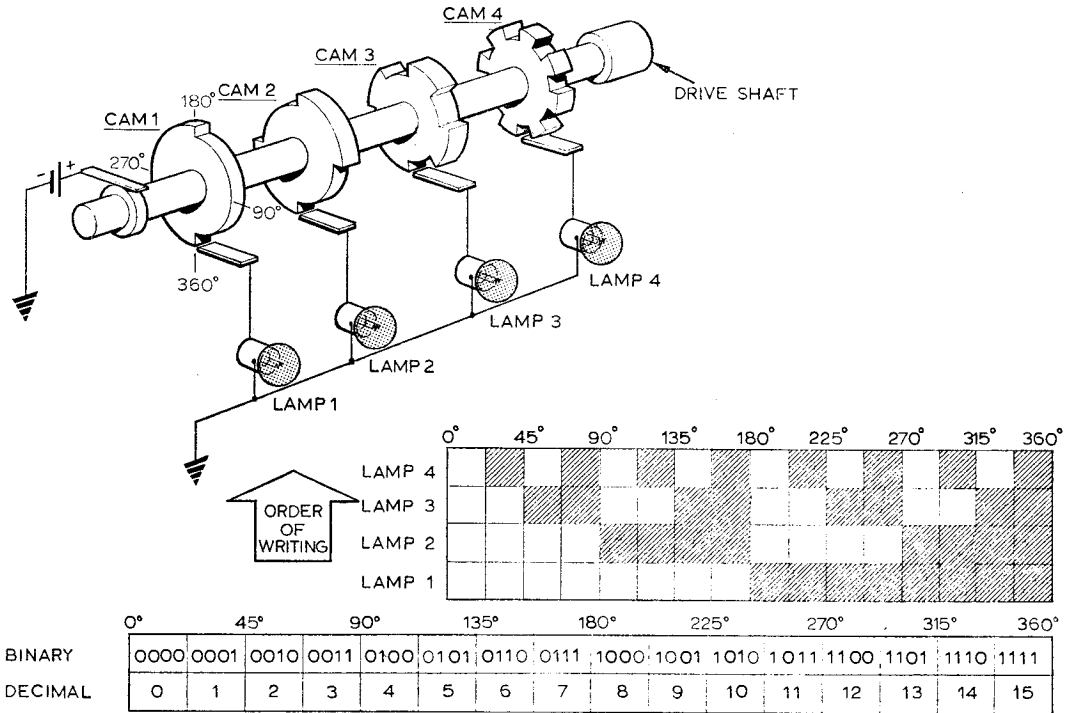


FIG 2. CONVERSION OF SHAFT ANGLE TO DIGITAL FORM, FOUR BINARY DIGITS

With 6 cams, 64 different sectors can be indicated. In fact with n number of cams, 2^n different sectors can be distinguished. One *shaft encoder*, as these devices are called, uses 11 cams providing 2048 sectors each approximately $10\frac{1}{2}$ minutes of arc wide (0.16° approximately).

It should be noted that the lamps and cams used in the above description are solely for the purpose of illustrating the basic principle of shaft-to-digit conversion. A practical shaft encoder,

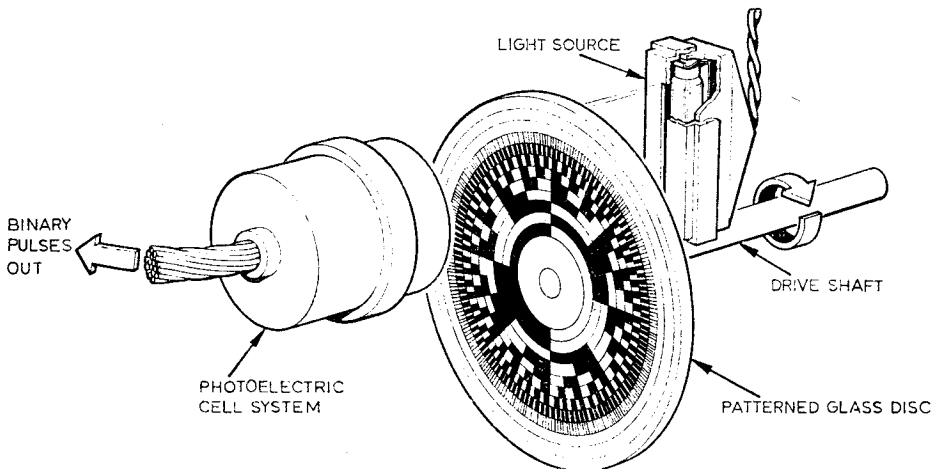


FIG 3. PRACTICAL PHOTOELECTRIC SHAFT ENCODER

although using the same basic principle, would be quite different. A practical device uses a thin glass disc mounted on the shaft and on this disc the required pattern has been photographically reproduced (Fig 3). Each circle of patterns corresponds to one mechanical cam. The black areas have the same purpose as the protrusions on the mechanical cam (binary digit 1) and the clear areas correspond to the undercuts on the mechanical cam (binary digit 0). A lamp produces a line of light at the surface of the disc and light passing through the clear areas illuminates a photoelectric cell system to produce the binary output pulses.

Parallel and Series Codes

The information from a shaft encoder is in *parallel* form, i.e. all the digits are presented *simultaneously*. To transmit this information as it stands would need a *separate* line or channel for each digit, and with a large binary number the cost would be prohibitive. So what we do is to transmit digits *serially*, one after the other. The information is therefore received serially and each digit must be stored until the whole number is available.

A simple method of serially transmitting the parallel information obtained from a four-cam shaft encoder is shown in Fig 4. Each cam contact producing a digit is connected to a stud; and a wiper arm, rotating at constant speed, makes contact with each stud in turn for a given period of time. Thus a lamp connected to the wiper arm indicates the information in serial form and a single line or channel to the receiver is all that is required for transmitting this item of digital information.

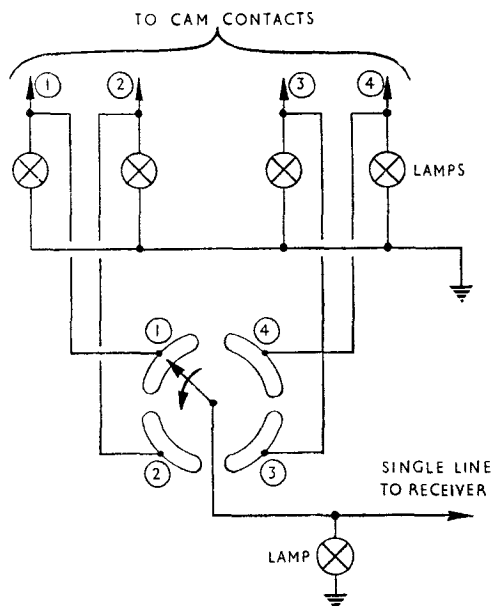


FIG 4. PARALLEL-TO-SERIES CONVERTER, SCHEMATIC

The Continuous Progression (CP) Code

If the shaft from which the numerical information is being obtained is continuously changing its position the binary information will be continuously varying. Where a change in value occurs *during* the serial transmission of the information, the use of the simple binary code can introduce large errors.

Let us suppose that the shaft position cannot change by more than one unit during the time it takes the wiper to sweep all the studs. If, during the time required to transmit the binary number 0111 (decimal 7) in serial form, the value changes to 1000 (8), then the number received may be 0110 (6), 0100 (4) or 0000 (0), depending upon the actual instant at which the change occurs. As the binary numbers increase in value, so the possible magnitude of error increases. A very large error could occur during a change from 0111111 (63) to 1000000 (64).

The continuous progression (CP) code reduces this error. The true binary code is rearranged by the cams so that the digital pattern for any two adjacent numbers differs by only *one* digital position. Fig 5 shows the CP code used in digitizing information from four cams. Thus, provided the value changes by only one unit during the serial transmission period, the error introduced cannot be greater than one unit.

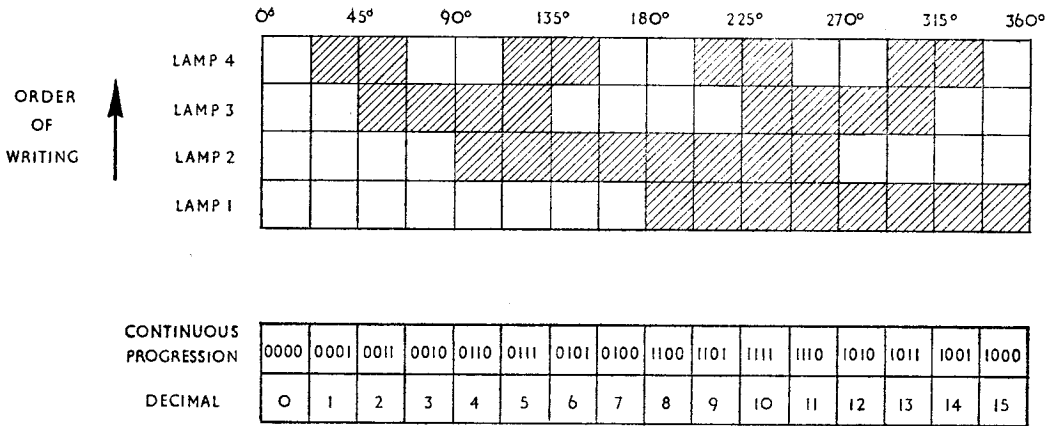


FIG 5. EXAMPLE OF CP CODE, FOUR-DIGIT SYSTEM

Principle of the Summation Coder

The other analogue-to-digital conversion with which we are concerned is that where we convert a d.c. analogue voltage input to its digital equivalent. The d.c. input may be an analogue of, say, radar range and we want to put this into digital form for transmission. One way in which this can be done is with the summation d.c.-to-digit coder, the outline of which is illustrated in Fig. 6.

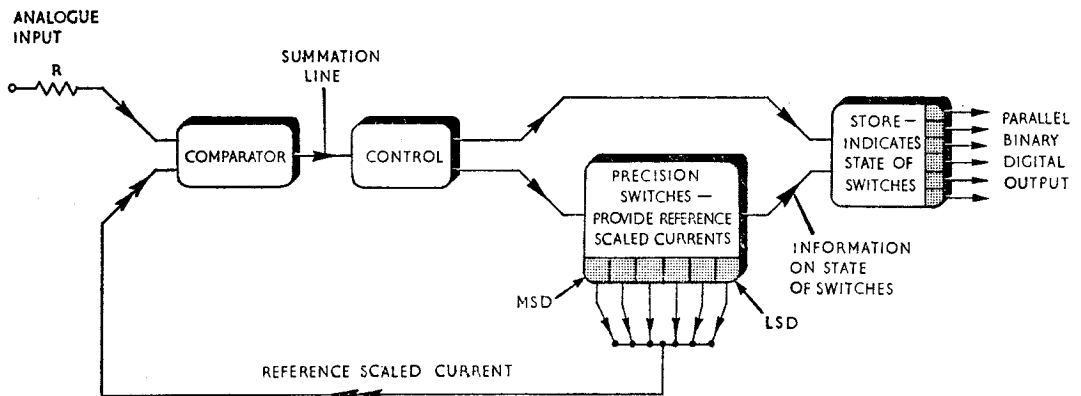


FIG 6. SIMPLIFIED BLOCK DIAGRAM OF DC-TO-DIGIT CODER

In this arrangement the principle of 'successive approximation' is used. The circuit includes a set of precision switches and the analogue input is compared with the output of each switch in succession. The circuit which compares the two inputs is a simple high-gain d.c. amplifier. It is used to detect when the current through R from the analogue input is equal and opposite to the current supplied by the switches. When balance is reached, the summation line output of the comparator is at zero volts and it remains at this value until the circuit operation is completed. In the meantime, the stores have noted the precision switch positions (1 for ON and 0 for OFF) so that the binary digital equivalent of the input analogue is available at the output.

The first comparison made is that between the input analogue and the most significant digit (m.s.d.) switch, i.e. the one which supplies the largest current. If the current due to the analogue input is greater than the switch current, the switch is left on (binary 1); if less, the switch is put off (binary 0). The other switches are then tried in turn and when the sequence is complete the stores contain information on the state of the switches and hence the digital equivalent of the analogue input signal.

Precision Switches

A suitable precision switch consists of two silicon diodes and a resistor R connected to a stabilized voltage reference source $+V$ as shown in Fig 7a. If point A is made positive relative to earth, the diode D_1 will be cut off and a current of $\frac{V}{R}$ amps will flow through D_2

and the indicating meter to earth. If point A is negative with respect to earth, current will flow in D_1 and the resulting voltage drop across R will cut D_2 off so that there is no current in the indicating meter. The switch is said to be ON in the first case and OFF in the second, the state of the switch being determined by the gating pulse.

If we make the reference voltage negative and reverse the diodes, as in Fig 7b, we have a precision switch which is cut on by a negative gating pulse.

The value of the current in the meter when the switch is on obviously depends upon the values of V (constant) and R . By having several switches with *different values* of R we can produce different values of current. Furthermore,

if the resistance values have a *binary relationship* to each other so also will the switch currents and we then have a set of binary digit precision switches.

Fig 8 shows a set of four binary digit precision switches. The resistance values, 32k, 64k, 128k and 256k, have the binary relationship 2^5 , 2^6 , 2^7 and 2^8 respectively. Thus the currents from the switches have a similar relationship, the most significant digit switch being S_1 . If S_1 is cut on by a positive gating pulse to A (all the other switches being off) the current through the meter has a value of $V/2^5$ mA. Similarly, if S_2 is on and all the other switches off, the current through the meter has a value of $V/2^6$ mA; and so on for the other switches. Hence, by means of this set of precision switches we can switch binary-scaled currents into or out of circuit as required. It is these currents that we use as a *scaled reference* against which to compare the input analogue current.

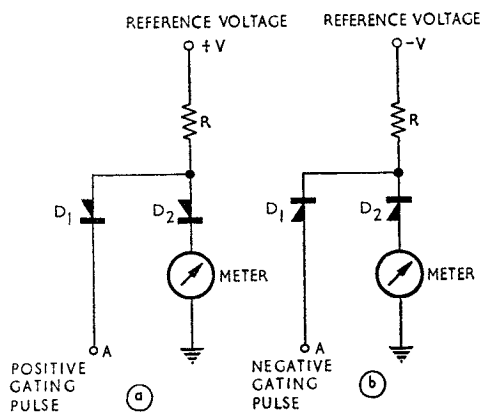


FIG 7. BASIC PRECISION SWITCHES

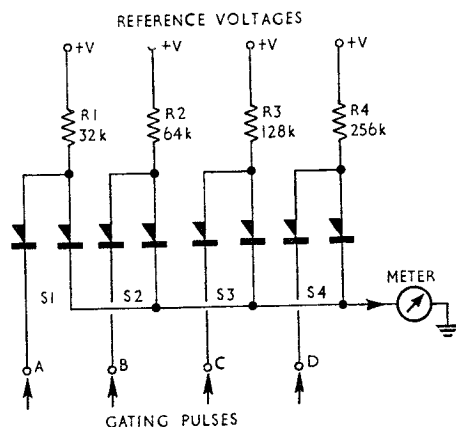


FIG 8. BINARY DIGIT PRECISION SWITCHES

DC-to-digit Coder

To encode a d.c. voltage into a binary number each digit is made to represent a certain magnitude of voltage on a binary scale. The least significant digit 01 may represent 1 volt, the next (10) 2 volts, the next (100) 4 volts, and so forth. In this way a 4-digit binary number could be used to represent a d.c. voltage of any whole number of volts between 0 and 15 volts: 1111 binary represents $8 + 4 + 2 + 1 = 15$ volts; 0101 represents 5 volts; 0000 represents 0 volts.

To indicate the polarity of a d.c. voltage the most significant digit is preceded by a 1 to indicate a positive voltage and a 0 to indicate a negative voltage. Thus, by using a 5-digit binary number a d.c. voltage of any whole number between -15 and $+15$ volts can be encoded. To encode a higher voltage with the same accuracy, a binary number having more digits must be used.

Table 1 shows the binary numbers which may be used to encode d.c. voltages between $+15$ and -15 volts.

Volts	Binary Number	Volts	Binary Number	Volts	Binary Number
+15	11111	+5	10101	— 5	01011
+14	11110	+4	10100	— 6	01010
+13	11101	+3	10011	— 7	01001
+12	11100	+2	10010	— 8	01000
+11	11011	+1	10001	— 9	00111
+10	11010	0	10000	—10	00110
+ 9	11001	—1	01111	—11	00101
+ 8	11000	—2	01110	—12	00100
+ 7	10111	—3	01101	—13	00011
+ 6	10110	—4	01100	—14	00010
				—15	00001

TABLE 1—BINARY-CODED NUMBERS

In this table the last four digits of the *positive* voltage values are actual binary numbers of the voltage they represent. However, for *negative* voltages this is not so. Thus -15 volts is represented by 0001, which is the binary number for 1; and -1 volt is represented by 1111, which is the binary number for 15. Hence to find the value of a negative voltage its binary number must be subtracted from 10000, i.e. from zero volts.

We now wish to combine the information given in the previous paragraphs on the summation coder, precision switches, and binary-coded numbers to produce an arrangement that will accept an input voltage analogue and convert it to a binary digit output. The schematic outline of such a system is shown in Fig 9 overleaf.

The purpose of this circuit is to cause the summation line eventually to take up a level of zero volts. At the same time, a digit pattern is produced in the stores and this pattern bears a direct relationship to the incoming analogue current.

Suppose that the input analogue voltage (representing, say, radar range) has a value of -13 volts. From Table 1 we see that the required binary-coded digital output is 00011. Because of the inverter, the voltage at point x has a value of $+13$ volts. Thus the current on the summation line due to the analogue input is $+\frac{13}{32} = +0.406$ mA.

The circuit action commences by P_1 pulse resetting all the digit stores so that they read 10000 (corresponding to zero volts input). This action switches off all the precision switches (S_1 is a negative switch similar to that shown in Fig 7b). Thus the only current on the summation line is the $+0.406$ mA due to the analogue input and this produces a positive-going output from the comparator to the gates.

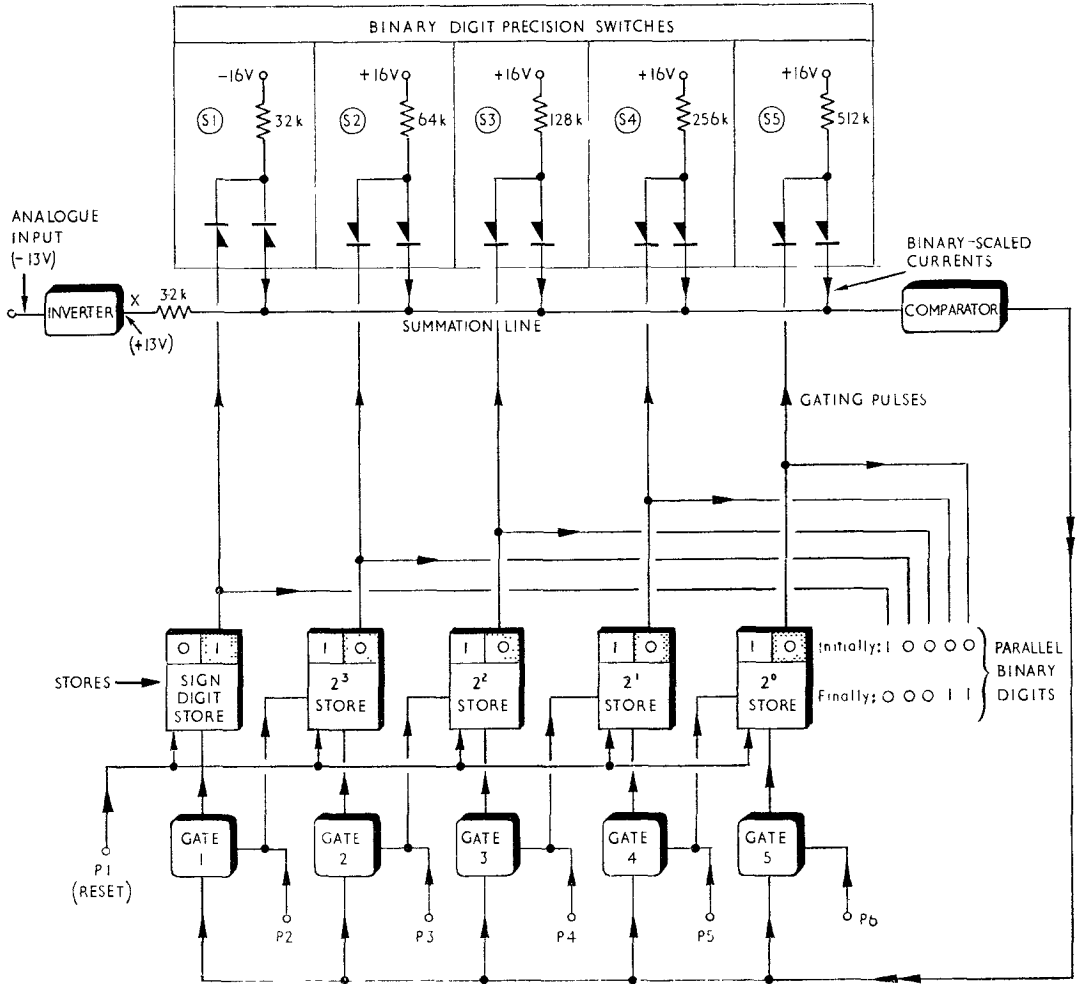


FIG 9. ARRANGEMENT FOR CONVERTING A DC VOLTAGE ANALOGUE TO A BINARY NUMBER

After a short interval P_2 pulse switches on gate 1 and the sign digit store is reset to '0'. This action switches on the precision switch S_1 which supplies a 'negative' current to the summation line of value $-\frac{16}{32} = -0.5$ mA. However, at the same time as P_2 pulse switched on gate 1 it also temporarily set the 2^3 digit store to '1'. This switches on precision switch S_2 which supplies a positive current to the summation line of value $+\frac{16}{64} = +0.25$ mA. Thus, during pulse P_2 time, S_1 and S_2 are both operative and the total current on the line, including the analogue current, is $+0.406 - 0.5 + 0.25 = +0.156$ mA. This positive current produces an output from the comparator to the gates.

The effect of the comparator output is that when P_3 pulse arrives, gate 2 opens and sets the 2^3 digit store to '0', switching off S_2 . At the same time, the 2^2 store is temporarily set to '1' and this brings S_3 into circuit to supply a positive current of $+\frac{16}{128} = +0.125$ mA to the summation line. With S_1 still on, S_2 now off and S_3 on, the total current on the line, including the

analogue current, is now $+ 0.406 - 0.5 + 0.125 = + 0.031$ mA. This again produces an output from the comparator.

Because we still have an output from the comparator, pulse P_4 opens gate 3, causing the 2^2 store to switch back to '0'. This switches off S_3 . At the same time, the P_4 pulse sets the 2^1 store to '1' and this brings S_4 into circuit to supply a positive current of $+\frac{16}{256} = + 0.063$

mA to the summation line. With S_1 still on, S_2 off, S_3 off and S_4 on, the total line current is now $+ 0.406 - 0.5 + 0.063 = -0.031$ mA. The input to the comparator is now negative, and with a zero or negative input there is *no output* from the comparator.

Thus, when P_5 pulse appears at gate 4 there is no output from the gate and the 2^1 store *remains* at '1', and S_4 is still operative. At the same time, P_5 pulse sets the 2^0 store to '1' and this brings S_5 into circuit to supply a positive current of $+\frac{16}{512} = + 0.031$ mA to the line.

We now have S_1 on, S_2 off, S_3 off, S_4 on and S_5 on, and the total line current, including the analogue current, is $+ 0.406 - 0.5 + 0.063 + 0.031 = 0$. The summation line is now at *zero* and there is *no further output* from the comparator.

Thus, when P_6 pulse appears at gate 5 there is no output from the gate and the 2^0 store remains at '1', leaving S_5 operative. *Balance* has now been reached and the digit stores read 00011, the required binary code for a d.c. analogue input of -13 volts.

This whole sequence of events takes place very rapidly and is repeated continuously to cater for any change in the input analogue value.

The output in this example is in parallel form. For baseband modulation of the link transmitter the digital data have to be in *serial* form. This is obtained by applying the binary output to a parallel-to-serial converter.

Summary

In this chapter we have considered ways in which shaft angle and d.c. voltage analogues of radar information may be converted into digital data form. The digital information is then used to modulate a microwave link transmitter to supply the required information to a distant control centre. Other forms of analogue-to-digital converters have been used in practice. However, the basic ideas are as described in this chapter.

RADAR DATA DISPLAYS

Introduction

In a modern air traffic control centre dealing with a large number of aircraft the information obtained from the radar concerning each target must be presented to the controllers in a way that can easily be interpreted. In addition to a 'raw' radar display showing all targets from the many radar sites, each controller requires displays which will give detailed information about the targets with which he is personally concerned.

To this end a variety of displays are available to the controllers including raw radar, synthetic radar, tabular and closed circuit television displays. Thus a controller's console could contain the displays shown in Fig 1. In this chapter we shall consider these displays.

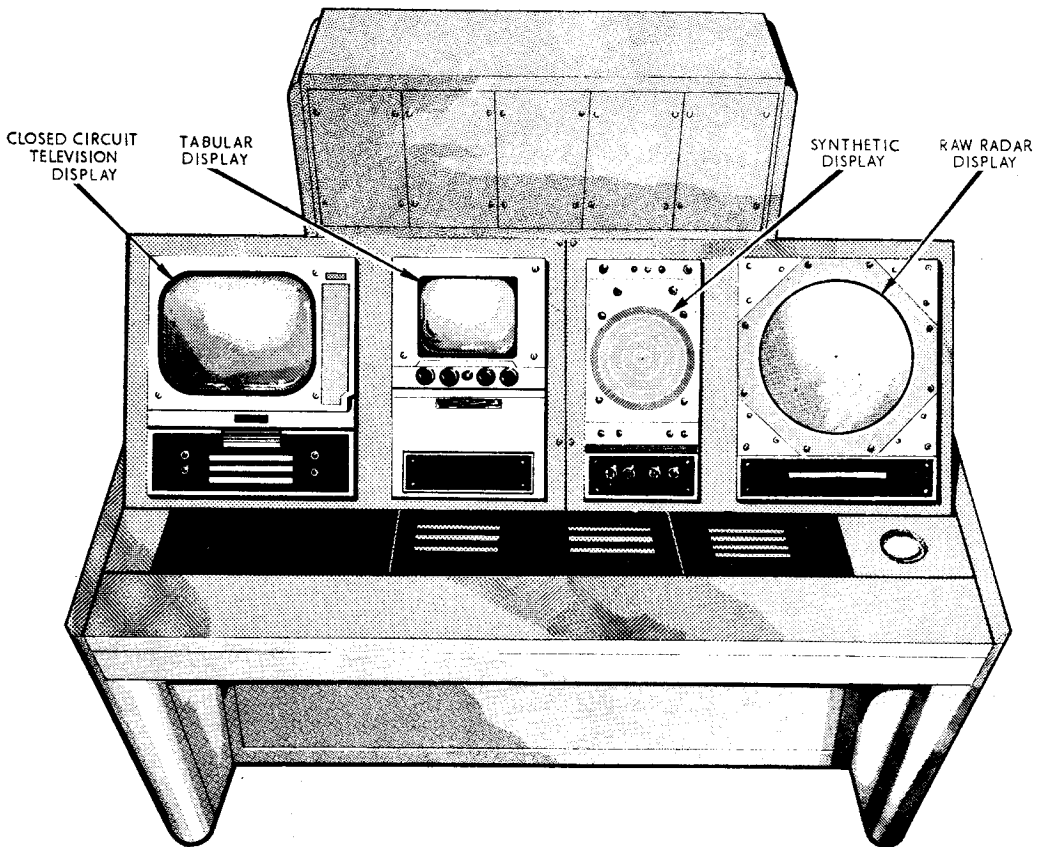


FIG 1. TYPICAL RADAR DISPLAY CONSOLE

Synthetic Display

In a modern air traffic control system, radar information from primary and secondary radars concerning the range, azimuth angle, height and identity of a large number of aircraft is fed, in digital form, into a digital computer. The range and azimuth angle, which together give the position of the aircraft, are changed into x and y cartesian co-ordinates and corrected so that

the aircraft position is *relative to the control centre position*. In this new form the positional data are passed to a digital store and up-dated on every aerial sweep. The computer calculates the aircraft track from successive positions, and this, together with height and identity data on each aircraft, is passed into another section of the computer store (Fig 2).

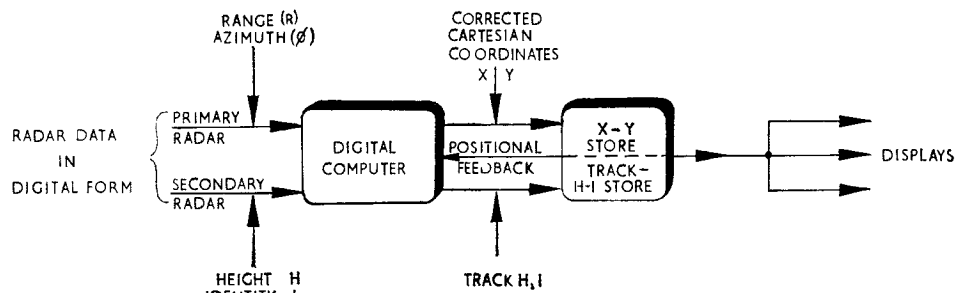


FIG 2. COMPUTER PROCESSING OF RADAR DATA

The outputs from these stores are fed to a *synthetic display*. This is a normal type of grid-modulated p.p.i. display fed with *processed* radar data. With a synthetic display the controller can select for display the targets in which he is interested. This is done by operating switches which send demands for the required data to the computer.

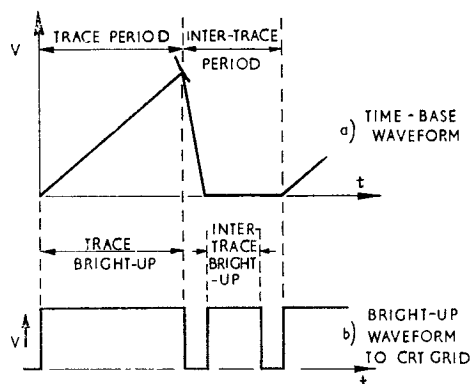


FIG 3. USE OF INTER-TRACE PERIOD FOR CHARACTER WRITING

As well as the target 'blips' displayed on the c.r.t. screen each target can be marked with identification symbols and characters. The c.r.t. has a dual fixed-coil deflection system: one produces the radial scan synchronized to the rotating radar aerials; the other is an *inter-trace* deflection system. This second deflection system operates during the inter-trace period of the main scan (Fig 3a) and moves the electron beam so that characters are written on the screen (see p216). To enable this to be done an inter-trace bright-up pulse is applied to the c.r.t. grid (Fig 3b).

These characters can be either in the form of *strobe markers*, controlled by the radar operator, or groups of letters and numerals, known as *alpha-numeric* characters, originated by a character generator controlled by the computer. A simplified diagram showing the connections of equipment for inter-trace marking is given in Fig 4. The flyback time of the main time-base generator

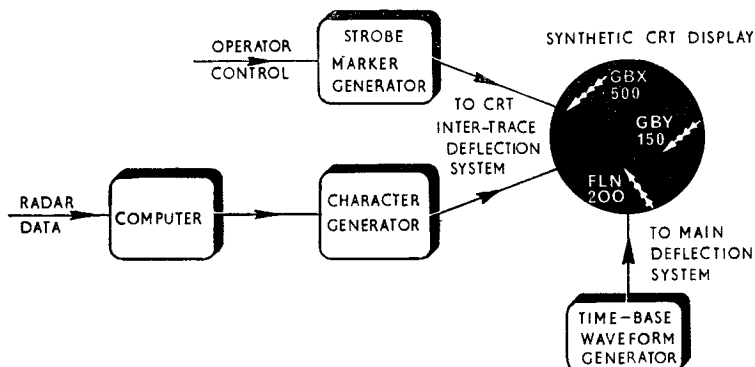
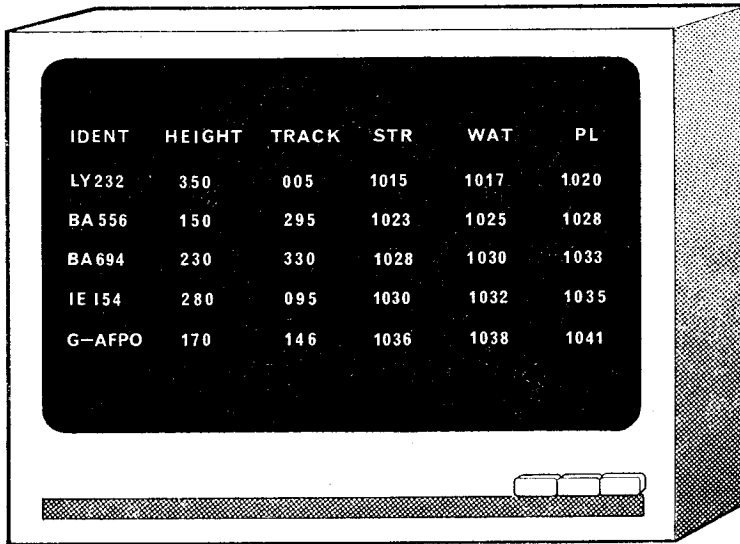


FIG 4. ARRANGEMENT FOR PRODUCTION OF SYNTHETIC DISPLAY

is made as short as possible so that the maximum number of alpha-numeric characters may be written. The data concerning track, height or identity held in the computer store are fed to the character generator which produces currents to deflect the c.r.t. beam and form the required character. Writing speeds of up to 50,000 characters per second are possible, each character being written in 20 microseconds by up to sixteen 1-microsecond deflections of the c.r.t. beam.

Tabular Displays

A tabular display is one in which target information such as identity, height, track and estimated time of arrival at reporting points is written on the face of a c.r.t. in alpha-numeric characters (Fig 5). The last three columns in Fig 5 indicate estimated times of arrival at various



IDENT	HEIGHT	TRACK	STR	WAT	PL
LY 232	350	005	1015	1017	1020
BA 556	150	295	1023	1025	1028
BA 694	230	330	1028	1030	1033
IE 154	280	095	1030	1032	1035
G-AFPO	170	146	1036	1038	1041

FIG 5. TYPICAL ALPHA-NUMERIC TABULAR DISPLAY

points. A tabular display is separate from the radar p.p.i. display and *does not contain target blips*. The data may be derived from a digital computer or from a keyboard similar to that of a typewriter.

In an automatic data processing system the information to be displayed is processed and passed to a digital store where it is kept up to date. The stored data are sequentially scanned and used to control an electronic character generator which provides the waveforms required to produce the characters on the display c.r.t. As the information is up-dated a single character on the display can be changed without altering the remaining characters.

The information is displayed on a c.r.t. screen as alpha-numeric characters in a series of horizontal lines. Each character is written by continuous movement of an electron beam through the intersections of a 'modulated' matrix. The picture can be read in normal daylight at a distance of three feet.

Typically, the equipment can write 50,000 characters per second, the c.r.t. afterglow being such that a steady picture results. Several displays can be fed from a control position, and the controller can select information concerning any target in which he is interested.

Video Maps

Video maps provide a means of associating a radar p.p.i. picture with geographical features such as cities, airfields, coastlines and mountains, and with reference systems such as the georef grid, reporting beacons and airway boundaries. The video map is mixed with the radar video

echoes fed to the p.p.i. c.r.t. and appears as lines traced on the c.r.t. screen. The controller can thus establish the position of a target with reference to important geographical and reference points.

The map to be displayed is photographed and a photo-negative slide is obtained on which the coast outline, airfields, etc. are transparent. Fig 6 illustrates a slide which shows a coastline, airfields, airway boundaries and aircraft reporting points.

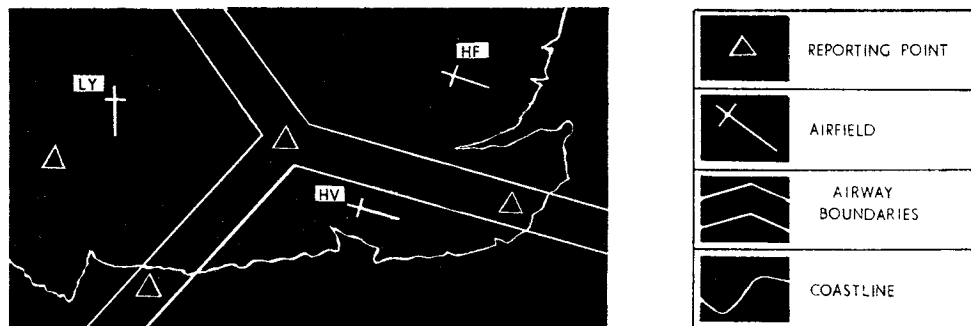


FIG 6. PHOTOGRAPHIC SLIDE FOR VIDEO MAP

The basic components of an equipment which produces a video map on the p.p.i. screen from a map slide are shown in Fig 7. The projection c.r.t. is a flat-faced tube giving a finely-focused spot of light. The screen is coated with a non-persistent luminous substance and the electromagnetic deflection coil is rotated round the neck of the tube by a power selsyn driven from the aerial head. The timebase, triggered by the radar master timing pulse, is fed to the deflection coil. Thus the spot of light on the projection c.r.t. screen moves radially out from the centre of the c.r.t. each time the radar transmits a pulse and moves round the face of the c.r.t. in synchronism with the rotating aerial.

A wide-angle optical lens focuses the spot of light on to the map slide on which the lines of information are transparent. The light which passes through the slide is then focused on to the light-sensitive cathode of a photo-electric multiplier. This converts the pulses of light into electrical energy. (See p346 Part 1B of these notes for the action of a photo-electric multiplier).

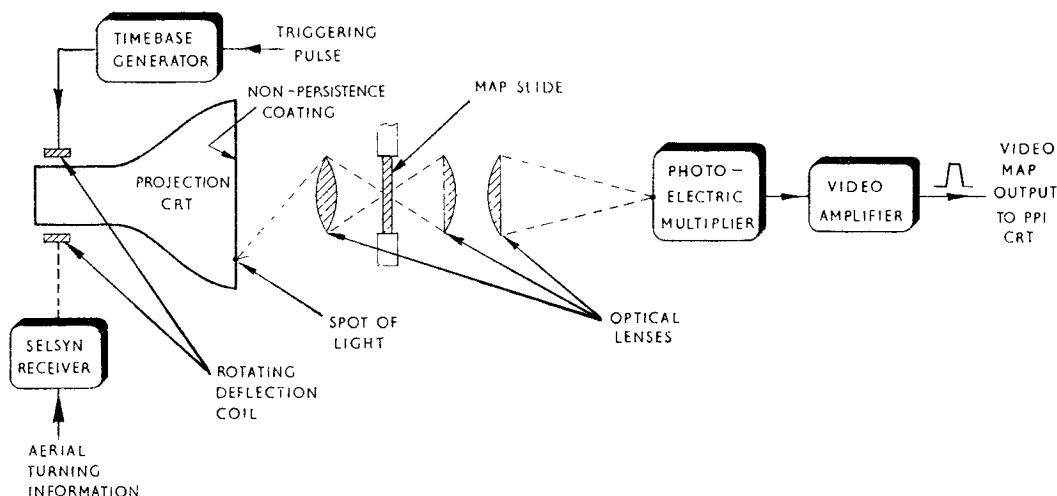


FIG 7. OUTLINE OF EQUIPMENT FOR PRODUCING A VIDEO MAP

The output from the photo-electric multiplier consists of pulses of current of varying amplitude and duration. These are amplified and limited in a video amplifier and then passed as constant amplitude pulses to the display p.p.i. The video map signal can be applied to several p.p.i. displays through matching cathode followers.

The video map produced from one slide will only be suitable for use on a single range-scale radar display. If a video map is required for a second range-scale, or for a different height zone, another map slide must be produced. This can be mounted alongside the other slide in front of the same projection c.r.t. but a separate optical system and photo-electric amplifier will be required.

The map slide is made from a master drawing and reduced to the required size when producing the negative. The slide is mounted in a carrier which must be accurately positioned relative to the lens. Usually two range markers are marked on the plate at 20 per cent and 80 per cent of maximum range and these can be used in setting up the map picture against a crystal-controlled range-marker generator.

Closed Circuit Television

Closed circuit television monitor screens are installed alongside the controller's radar p.p.i. and tabular displays. Flight information is initiated remotely by an assistant who writes on edge-lit perspex boards in chinagraph pencil. A television camera mounted behind each board produces a high-definition picture of the written information for each controller.

The perspex boards are mounted in a group and the output from each camera is fed into vision distribution amplifiers which provide twenty or more outputs from each camera.

These outputs are applied via vision switching units to the controller's monitors (Fig 8). Each switching unit is remotely operated by push-button selectors at the controller's viewing positions. Thus the controller can select the picture and hence the flight information he desires.

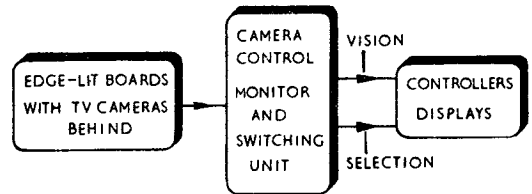


FIG 8. CLOSED CIRCUIT TELEVISION VIEWING

Summary

This section has served as an *introduction* to some of the modern methods used to collect, transmit, process, and display radar information. We have seen the need for microwave radar data links and the outline of possible systems. The methods of converting analogue values into digital form have been discussed but in this field, as with other subjects covered in this section, rapid development is taking place and the trend is towards more accurate and compact converters.

Data processing plays an important part in the handling and display of radar information. The advantages of fully automatic data processing as an integral part of a radar control system have been mentioned. The additional information obtainable from secondary radars is of considerable importance in the control of aircraft, as are the methods used to display this information to the controllers.

SECTION 9

SOME MODERN DEVELOPMENTS IN RADAR

Chapter		Page
1	Developments in Microwave Semiconductor Devices	543
2	Developments in Microwave Valves, Ferrite Devices and Delay Lines ...	569
3	Developments in Airborne Radar Systems	595
4	Primary Ground Radar Developments	609
5	Secondary Surveillance Radar	627

CHAPTER 1

DEVELOPMENTS IN MICROWAVE SEMICONDUCTOR DEVICES

Introduction

Research into ways in which radar performance may be improved is a continuous process. Over the past few years this research has resulted in the development of new techniques and new components covering the whole field of radar. Some of these recent developments have already been introduced into RAF equipment and others will undoubtedly be applied in the future. It is, therefore, important that we should know something about them. The developments of particular interest include:

- a.* New solid state semiconductor devices.
- b.* The introduction of thin-film and integrated circuits (microelectronics).
- c.* Improvements in electronically-scanned aerial arrays (phased-array radars).
- d.* Advances in airborne radar systems, including the use of sideways-looking radars and terrain-following radars.
- e.* Advances in ground radar systems, including the use of pulse-compression techniques and improved tracking systems.
- f.* Developments in secondary radar systems.
- g.* The increasing use of computers for data-handling and processing of radar information, leading gradually to the concept of the automatic radar.

In this Section we shall discuss some of these subjects in general terms. This will merely *introduce* these subjects so that we may have some idea of what to expect in future equipment. We shall start in this Chapter by considering developments in microwave semiconductor devices and in microelectronics.

Microwave Semiconductor Devices

Semiconductor theory is considered briefly in AP3302 Part 1B; semiconductor diodes are dealt with in p240 and transistors in p289. We have also previously considered some microwave semiconductor devices in p341 of this book; the devices dealt with there are the point-contact crystal diode, the tunnel diode, the PIN diode, and the varactor diode. Many other microwave semiconductor devices are now available—charge-storage (step-recovery) diodes, backward diodes, “hot electron” diodes, avalanche diodes, and Gunn-effect devices.

The active devices necessary at microwave frequencies can be split into:

- a.* High-power devices for the output stages of radar transmitters.
- b.* Medium-power and low-power devices for radar receivers, the initial transmitter stages, microwave links, and computers.

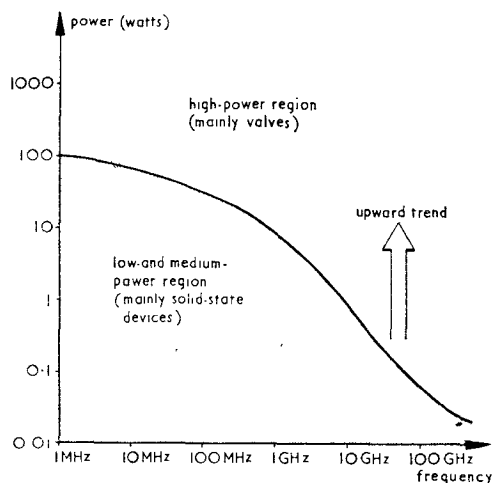


FIG 1. POWER AREAS FOR VALVES AND SOLID-STATE DEVICES

resistance portion of their current-voltage characteristic (e.g. tunnel diode and avalanche diode oscillators). More recently, Gunn-effect devices have been developed as microwave oscillators.

For *amplification* at microwave frequencies we have at the moment a choice of transistors, tunnel diodes, parametric amplifiers, masers, and Gunn-effect devices. Transistors have now been developed to operate satisfactorily in amplifier circuits up to about 3GHz. At higher frequencies it appears that the tunnel diode amplifier will continue to be used except where low noise is a requirement. At the moment the choice for high-frequency, low-noise amplification lies between the parametric amplifier and the maser; the latter, because of its complexity, will be used only for the large installations where ultra-low noise levels are vital. Gunn-effect amplifiers, although showing promise, have not yet been adequately developed.

The point-contact crystal diode is expected to continue to be used extensively as a microwave *mixer*. But newer devices, including the backward diode and the "hot electron" diode, are being developed to provide improved sensitivity and burnout protection.

Microwave semiconductor *switches* may be used as part of a radar TR switch, as switched attenuators, and as controlled modulators. PIN and varactor diodes are the main contenders here.

Frequency and Power Limitations in Transistors

The development of solid-state devices to give improved performance at higher frequencies and higher power levels has followed almost the same road as that taken earlier by valves. It will be remembered that the performance of conventional valves at high frequencies is limited by *transit time* effects. This may be reduced by very close spacing between the electrodes, as in the disc-seal valve. But this close spacing increases inter-electrode capacitances and reduces the gain at high frequencies. A reduction in the *area* of the electrodes was the next step but, although this reduces the inter-electrode capacitances, it also *limits the power* that the valve is able to

The high-power region is, for all practical purposes, catered for by valves (magnetron, klystron, travelling-wave tube, and amplatron). The other region is gradually being taken over by solid-state devices, where their greater reliability and lower d.c. power requirements can be used to advantage. This is illustrated in Fig 1, which also shows that solid-state devices are progressively invading the high-power region.

In those power areas where solid-state devices are used, the devices can have several roles: as generators, amplifiers, mixers, or switches.

Microwave *generation* may be provided by a microwave transistor oscillator; this may be followed by varactor, or step-recovery diode, frequency-multiplier chains to provide outputs at frequencies greater than 100GHz. Other sources of microwave power include semiconductor diodes operated over the negative-

dissipate. Thus, at frequencies greater than about 1GHz the conventional valve introduces so many practical difficulties that, to overcome them, klystrons and travelling-wave tubes were developed.

The semiconductor story is very similar. The main factors limiting performance at high frequencies are transit time effects and junction capacitances. The frequency response of normal transistors is limited by the transit time of the carriers across the base region. In the newer transistors, transit time effects are reduced by using very thin base regions (less than 0.0001 in). But this increases the junction capacitance and limits the transistor frequency response. The increase in junction capacitance can be offset by using small junction *areas* but here, as with the valve, the power that the device is capable of dissipating is then reduced.

By using modern manufacturing techniques some of these limitations are being overcome, and microwave transistors, capable of amplification at reasonable power levels in the 3GHz band, are now available. Some of these new techniques are described in the following paragraphs. However, transit time effects still set the limit to the frequency at which normal transistors may be used. At frequencies above a few GHz other semiconductor devices are being investigated.

Microwave Transistors

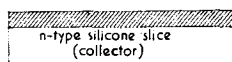
It has become common practice to refer to an n-p-n or p-n-p transistor as a *bipolar* transistor to distinguish it from another type of transistor which has been developed—the *field-effect transistor* (see later). The term bipolar is used because the current in n-p-n or p-n-p transistors is due to a simultaneous movement of *both* positive and negative charge carriers (holes and electrons). In an n-p-n transistor the majority-carrier electrons from the n-type emitter become *minority carriers* in the p-type base. Similarly, in a p-n-p transistor the holes become the minority carriers in the n-type base. The *transit time* is the time taken for the *minority carriers* to pass through the base region and is relatively much longer than that in valves. The minority carriers in a transistor flow through the base region to the collector by a process of *diffusion* under the influence of a very weak electric field. In contrast, the electrons in a valve are moving rapidly in a vacuum towards the anode under the influence of a large electric field.

The time taken for the minority carriers to cross the base region depends upon the *width* of the base region and also upon the “*mobility*” of the charge carriers—how quickly they move under diffusion conditions. The latter depends upon the semiconductor material used in the construction of the transistor. Germanium has a higher charge carrier mobility than silicon. Thus, for the same base width, the transit time in germanium transistors is less than that for silicon types. Germanium transistors are therefore able to operate at inherently *higher frequencies* than silicon types. However, advantage can be taken of this fact only at low power levels because of the temperature limitations of germanium.

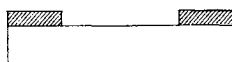
The other factor affecting the high-frequency performance of bipolar transistors is the *width of the base region*. For a typical low-frequency germanium transistor the base width may be 0.0005 in and this may give an upper operating frequency limit of, say, 12MHz. A germanium transistor designed to operate at 2GHz may require a base width of the order of 0.00005 in. Thus, *very thin bases* are the first requirement of high-frequency bipolar transistors.

Very precise and intricate manufacturing techniques are needed to produce such very thin bases. The whole device is extremely small, and highly-accurate photographic masks must be produced to define the pin-head-size areas. One of the most successful high-frequency bipolar transistors to date is the silicon planar diffused type. The basic steps in the production process are illustrated in Fig 2.

1. Surface oxidized to act as insulator.



2. Base 'window' etched out by photo-chemical process.



3. P-type base diffused in.



4. Surface re-oxidized.



5. Emitter window etched out.



6. N-type emitter diffused in.



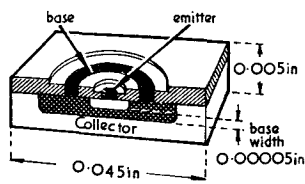
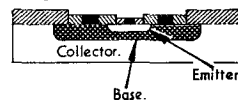
7. Surface re-oxidized.



8. Contact windows for base and emitter etched out.



9. Contacts made to active regions through windows.



- n-type semiconductor.
- p-type semiconductor.
- silicon oxide (insulator).
- metal contacts.

FIG 2. STEPS IN PRODUCTION OF SILICON PLANAR DIFFUSED TRANSISTOR

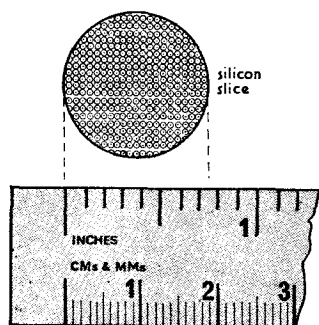


FIG 3. SIZE OF LOW-POWER MICROWAVE PLANAR DIFFUSED TRANSISTORS

Perhaps the most striking thing about modern high-frequency transistors—although not their most important quality—is their very small *size*. The dimensions are indicated in Fig 2 but to give meaning to such dimensions it is worth noting that over 300 planar transistor elements can be produced on a silicon slice of less than 1 in diameter (Fig 3).

As the frequency at which transistors are designed to operate is increased, the dimensions of the transistor *decrease*. Thus, the power that the transistor is capable of dissipating also decreases. To provide higher power outputs we must increase the collector and emitter junction *areas*. If this is done directly the junction capacitances increase and this limits the transistor frequency response.

However, in practice, transistor action is confined to an area adjacent to the *periphery* of the emitter. Therefore to increase the *effective* area a number of long, narrow emitter 'fingers' are used, with base contacts between them, as shown in Fig 4. This multi-stripe construction allows large collector currents to flow, the resulting heat being dissipated over a relatively large area. In general, the greater the number of stripes on one structure, the higher is the power output of the device. Transistors with up to 65 stripes have been developed.

Rather than concentrate on producing a high power output from a single device, the

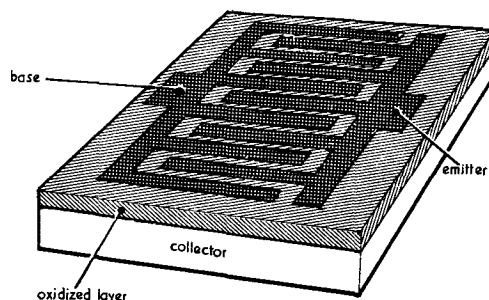


FIG 4. USE OF INTERDIGITAL STRUCTURE FOR POWER TRANSISTORS

alternative approach of connecting a number of transistors in parallel may be used. It has been found possible at the lower frequencies to connect up to 20 transistors in a parallel/push-pull arrangement to give a c.w. power output of 500W. This approach may well be attempted at the higher frequencies also.

The development work carried out on bipolar transistors has resulted in great improvements in performance at the higher frequencies. At the moment, the maximum frequency of *oscillation* is about 11GHz for germanium transistors and 9GHz for silicon types. The maximum operating frequency as *amplifiers* is about 3GHz, but experimental results increasing this figure to 6GHz have already been reported.

The current state of high-frequency transistors in terms of available power output, gain, and noise figure is indicated in Table 1.

Frequency	Power Output	Gain	Noise Figure
500MHz	20W	20db	3db
1GHz	5W (10W-under development)	16db	3.5db
2GHz	1.8W (5W-under development)	10db (16db-under development)	6db

TABLE 1. PERFORMANCE FIGURES FOR PRESENT HIGH-FREQUENCY TRANSISTORS

Frequency Multiplication in Transistors

Transistor oscillators or amplifiers are often used to drive a varactor, or step-recovery diode, frequency multiplier. In some microwave transistors the collector-base 'diode' of the transistor itself may act as a frequency multiplier. Under large-signal conditions the variation in the collector-base capacitance gives good conversion efficiency at the second and third harmonics of the input signal.

For example, one transistor amplifier operating at its fundamental frequency of 2GHz delivered 250mW of power with 9.5db gain. As a doubler it delivered 155mW (with 50mW input) for a conversion gain of 7db at 4GHz. The efficiency of the doubler was 20%. As a trebler, the efficiency would be lower.

The single transistor operating as an amplifier-multiplier has the advantage of simplicity compared to the transistor-varactor multiplier.

Metal-base Transistor

If the minority carrier transit time in the base region could be eliminated this would greatly improve the high-frequency performance of the transistor. This is being attempted in the metal-base transistor, in which the usual n-type or p-type base region is replaced by a thin *metal film* of molybdenum. This has three effects: the 'energy gap' between the emitter and the base is increased so that electrons from an n-type emitter are injected into the base region with *high energy* (see p355 on energy bands); the base resistance is reduced; and there are no minority carriers. We have, in effect, two 'hot carrier' diodes connected back-to-back, the 'hot carrier' label being used to describe the injection of high-energy electrons from the emitter. Electrons are injected over the emitter-base barrier and collected after crossing the reverse-biased base-collector junction. Control is effected, as usual, by the base voltage. The current in a metal-base transistor is no longer diffusion-limited, because there are *no minority carriers* in the base region.

In fact, the emitter current is more like the thermionic emission current in a valve, so that the mobility of electrons through the base is high and transit time effects are negligible. In addition, the base resistance of a metal-base transistor is low and this gives a further improvement in high-frequency performance.

A metal-base transistor amplifier using strip-line microwave circuit components has been developed to give 20db of gain at 10GHz. Further work is proceeding to reduce the noise figure for the device to make it suitable as an r.f. amplifier in microwave receivers. Manufacturing techniques have not yet been perfected but there is every indication that it will be possible to produce metal-base transistors for use with waveguide circuitry to provide amplifiers and oscillators in the 50GHz to 100GHz range.

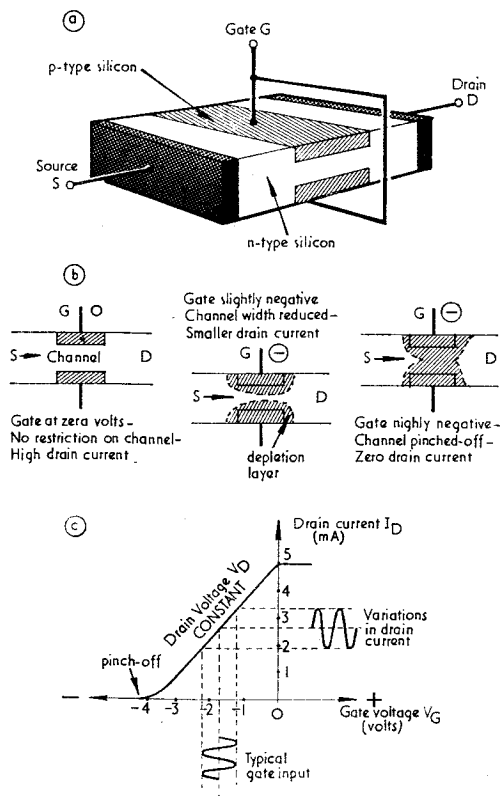


FIG. 5.

BASIC FET, CONSTRUCTION AND OPERATION

voltage and also upon the resistance of the n-type conducting channel between source and drain. This resistance is determined by the effective *width* of the channel between the p-type gate regions. If the gate electrode is made *negative* with respect to the source, *depletion layers* are formed between each p-type gate region and the n-type channel, as in any reverse-biased p-n junction (Fig 5b). Since a depletion layer acts as an insulating layer, having been depleted of charge carriers, the n-type conducting channel between source and drain becomes *narrower*. Its resistance therefore increases and the drain current decreases. By making the gate more negative, the n-type conducting channel becomes narrower still and the drain current reduces further. Eventually, if the gate electrode is made sufficiently negative with respect to the source, the two depletion layers join and there is now no conducting channel between source and drain.

Field Effect Transistors

The field effect transistor (f.e.t.) is a 'majority carrier' type of transistor which, because it has no minority carriers and hence negligible transit time effects, is capable of operating at very high frequencies.

The schematic outline of a basic f.e.t. is illustrated in Fig 5a. It consists of a bar of n-type silicon into the sides of which two regions of p-type semiconductor have been diffused. The p-type regions are normally connected together externally and form the control electrode known as the *gate*. Ohmic contacts at each end of the bar form the other two electrodes, the *source* and the *drain*. The source is equivalent to the emitter in a bipolar transistor; the gate is equivalent to the base; and the drain is equivalent to the collector.

It is clear from Fig 5 that a conducting n-type channel exists between source and drain, provided the gate is on open circuit. If we now make the drain positive with respect to the source, majority carriers (electrons in this case) flow from the source to the drain, channelling between the p-type gate regions. The magnitude of the drain current depends upon the drain

The drain current is zero. This effect, similar to cut-off in a valve, is known as '*pinch-off*'. A typical characteristic illustrating these points is shown in Fig 5c.

From what has been said it may be seen that by applying a signal voltage to the gate electrode, the drain current is caused to vary in a similar manner. By suitable choice of load resistor, the output voltage variations produced by the drain current can be much larger than the variations in signal voltage applied to the gate. Thus, *amplification* may be obtained.

FETs may also be constructed using a bar of p-type silicon into which n-type gate regions have been diffused. Provided all the polarities of voltages are reversed the action is similar to that described above.

The type of f.e.t. described in the preceding paragraphs was first introduced about 1953. Advances in technology since that date have brought considerable improvements in manufacturing techniques, and cheaper and more efficient f.e.t.s are now available. Typical of these is the n-channel *insulated-gate* f.e.t., illustrated in Fig 6a. This type of f.e.t. is also known as a *metal-oxide-semiconductor-transistor* (m.o.s.t.). The symbols for a m.o.s.t. are shown in Fig 6b.

If a voltage is applied between source and drain, leaving the gate on open circuit, only a very small leakage current flows between source and drain, because we have two p-n junctions back-to-back. Thus, with an insulated-gate f.e.t., pinch-off occurs at a gate voltage of *zero*. If now we apply a *positive* voltage to the gate terminal a *negative* charge is *induced* on the surface of the silicon under the oxide layer and this forms an n-type conducting channel between source and drain (Fig 6c). A current now flows between source and drain, the magnitude of the current being controlled by the voltage of the gate electrode. There is *no current* in the gate electrode; so the f.e.t. is a *voltage-operated device*, like a thermionic valve. If the gate voltage is increased, the *depth* of the induced conducting channel increases, allowing a larger drain current to flow. Thus the gate voltage controls or '*modulates*' the drain current, a *small change* in gate voltage producing a *large change* in drain current.

To operate as an amplifier a d.c. voltage is applied between the source and drain terminals through a load resistor. With an n-m.o.s.t. the drain terminal is made positive relative to the source, which is normally earthed, and the input signal is applied to the gate in series with a suitable positive bias. For a p-m.o.s.t. these polarities are reversed. As the gate voltage varies in sympathy with the signal, the depth of the induced conducting channel varies, and the drain current follows this variation to produce an amplified signal voltage across the load resistor.

Since the gate terminal is formed on a layer of silicon oxide (an insulator), the input impedance of the device is very high—of the order of several thousand megohms. It is similar to the

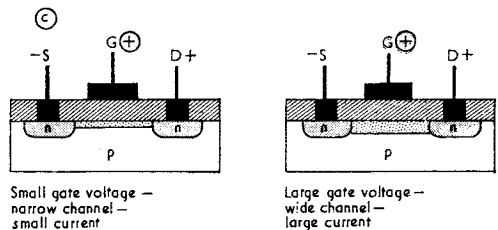
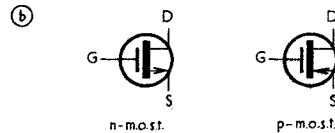
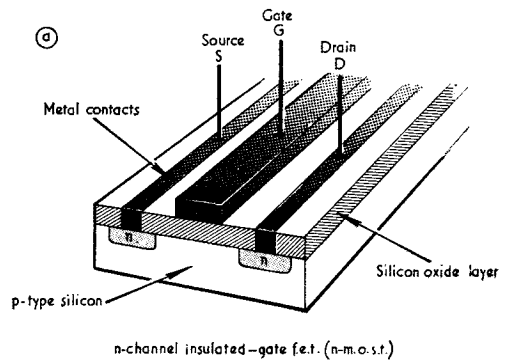


FIG 6. BASIC N-CHANNEL INSULATED-GATE FET (N-MOST)

thermionic valve in this respect, and in contrast to the bipolar transistor with its low input impedance. The f.e.t. also has a low noise figure and should prove effective for high impedance, low noise, input stages. In effect, the f.e.t. combines the relative advantages of valves and transistors.

We have previously noted that the f.e.t. is a majority carrier transistor. In the n-m.o.s.t. the charge carriers are electrons. There are no minority carriers, so that transit time effects are negligible. The input signal merely varies the *width* of the conducting channel and *does not vary the rate of travel of the current carriers*. In addition the f.e.t. has no base region so that two of the factors previously affecting the cut-off frequency in the bipolar transistor—base width and base resistance—are eliminated. The only factors limiting the frequency of operation are certain shunting capacitances and these are less than for the bipolar transistor. Thus, inherently, the f.e.t. is capable of operating at very high frequencies in the microwave region. Several production problems have yet to be overcome before this high potential is fully realized, but already f.e.t. amplifiers are available for operation at frequencies in excess of 500MHz with gains of 12db and noise factors of 4db. These figures will almost certainly be improved.

Other types of f.e.t. operating in a different manner are available. However, all f.e.t.s work on the principle that a conduction channel between source and drain can be “modulated” by an input signal applied to the gate electrode, the resulting drain current producing an amplified signal voltage across the load.

Varactor Diodes

The varactor diode is considered in p344 of this book. There we learned that the *variable capacitance* properties of this diode may be used in microwave switching, frequency multiplication, and parametric amplification.

In common with other semiconductor devices, progress has been made in producing varactor diodes capable of operating at higher frequencies and at higher power levels. For a varactor diode to operate efficiently, the reactance of the junction capacitance should be *much greater* than the series resistance of the diode. As the frequency applied to the diode is increased so the capacitive reactance *decreases*. Eventually, if the frequency is increased sufficiently, the capacitive reactance will fall to a value *equal* to the diode series resistance. The frequency at which this happens is known as the *cut-off frequency* of the varactor. Note, however, that the varactor efficiency is decreasing progressively as the frequency increases, and long before the cut-off frequency is reached the diode is useless as a varactor.

To improve the efficiency of varactor diodes at higher frequencies, the cut-off frequency must be increased. The majority of varactor diodes to date have used silicon as the semiconductor and with the latest diffusion techniques (as described for the bipolar transistor on p.546) very small junction capacitances and low series resistances can be obtained. Cut-off frequencies greater than 100GHz result, allowing efficient operation of the varactor up to about 10GHz.

The semiconductor *material* used also has an influence on the cut-off frequency. For high cut-off frequencies semiconductor materials with high charge-carrier mobility are required. Gallium arsenide is such a material and diffused gallium arsenide varactor diodes have been designed with cut-off frequencies of 350GHz. For still higher cut-off frequencies very small junction capacitances must be produced (of the order of 0.2pF). This can be achieved by using a *point-contact* gallium arsenide crystal diode. Experimental results with such a diode

have indicated the possibility of cut-off frequencies of the order of 1,000GHz. If this figure is achieved in practice, varactor diodes capable of operating in the 100GHz band will be realized.

The requirements of a varactor diode depend upon whether it is to be used in a parametric amplifier or in a frequency multiplier. For optimum performance in a parametric amplifier the varactor diode should have a high cut-off frequency and a large variation of capacitance with voltage (Fig 7). For harmonic generation the diode must also have a high breakdown voltage (often as high as 100V), a high power-conversion efficiency, and good power-handling capacity.

The *efficiency* of a varactor diode as a frequency multiplier depends upon the harmonic being generated. As a *doubler*, efficiencies of the order of 75% have been obtained at microwave frequencies. This reduces to about 30% for a *trebler* and is lower still for higher-order harmonics. (The harmonic amplitude decreases as $n^{\frac{1}{2}}$ for a varactor diode, where n is the harmonic order).

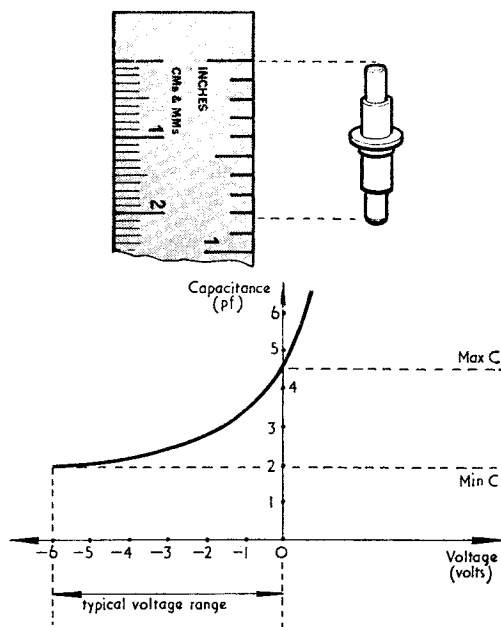


FIG 7. VARIATION OF CAPACITANCE WITH VOLTAGE FOR A TYPICAL VARACTOR DIODE

In the conventional frequency multiplier, only the fundamental and the desired harmonic currents are allowed to flow in the varactor.

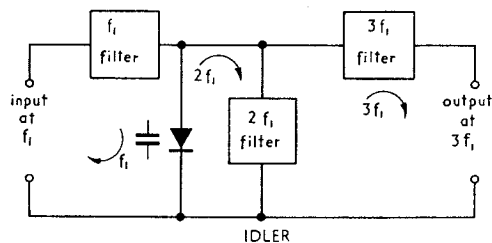


FIG 8. USE OF AN IDLER CIRCUIT TO IMPROVE TREBLER EFFICIENCY

The efficiency of the trebler and higher-order multipliers can be improved by using 'idler' circuits in which other selected harmonic currents flow. For example, in a trebler, an 'idler' circuit tuned to the *second harmonic* may be inserted as shown in Fig 8. By using an idler circuit the efficiency of the trebler may be increased from 30% to about 60%, but this increase in efficiency is obtained only at the expense of increased circuit complexity. This applies particularly for the higher-order harmonics; to produce an x6 multiplier, idlers at the second, third, fourth, and fifth harmonics would be necessary. In such cases it may be just as efficient, and certainly less complex, to use *two stages* (x2x3) without idlers.

The *power capability* of varactor diodes used as frequency multipliers is also constantly improving. Present power outputs range from 100W at 100MHz to 0.1W at 100GHz. Bandwidth is usually about 1% of the operating frequency to allow for frequency modulation of the driver or electronic tuning of the oscillator.

The overall improvements in varactor diode frequency multipliers are indicated in Table 2, which shows figures for typical multiplier chains. Note particularly the high efficiency of c compared to b.

As a step towards *integration* of microwave circuits, a composite varactor consisting of four

Frequency In	Multiplication	Frequency Out	Power In	Power Out	Efficiency (Year)
a. 125MHz	x72 (x3x2x2x2x3)	9GHz	10W	150mW	1½% (1960)
b. 400MHz	x16 (x2x2x2x2)	6.4GHz	8W	300mW	4% (1961)
c. 625MHz	x16 (x2x2x4)	9.9GHz	2W	250mW	30% (1963)
d. 333MHz	x3	1GHz	40W	35W	65% (1965)
e. 4GHz	x3	12GHz	2W	1.2W	65% (1966)

TABLE 2. CHARACTERISTICS OF VARACTOR DIODE MULTIPLIERS

varactor wafers, series-connected in a single package, has been produced (Fig 9). When tested in a 4 to 12GHz trebler (see Table 2e) the composite varactor gave an 18-fold increase in power output compared to a single junction. Since *n* wafers in series have *n* times the resistance and $\frac{1}{n}$ times the capacitance of a single wafer, the cut-off frequency remains the same. With series-connected varactors, the breakdown voltage is greatly increased and the efficiency is high, even for high-order multiplication.

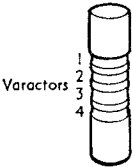


FIG 9. SERIES-CONNECTED VARACTORS

Charge-storage (Step-recovery) Diodes

Most varactor-diode frequency multipliers use the rectifying properties of the diode itself to provide the required operating bias, i.e. the drive voltage is made sufficiently large to drive the diode well into the forward-conducting region. However, there is a disadvantage in this: if a large rectified current is allowed to flow, the diode losses are increased and the efficiency as a multiplier is reduced.

During the development of the varactor diode for operation at higher frequencies it was noted that if the frequency is sufficiently high, as the polarity of the drive voltage reverses, large numbers of minority charge carriers tend to be *drawn back* across the junction depletion layer *before they have time to recombine with holes* (or electrons). As a result, the rectified current is reduced. This effect is known as *charge storage*. It is now known that charge storage effects provide better frequency-multiplier performance than the conventional depletion layer mechanism of the varactor diode. Because of this, special frequency-multiplier diodes *using* the charge-storage technique have been developed. The charge-storage (or step-recovery) diode is constructed such that the barrier confines injected minority carriers to the vicinity of the barrier region. We

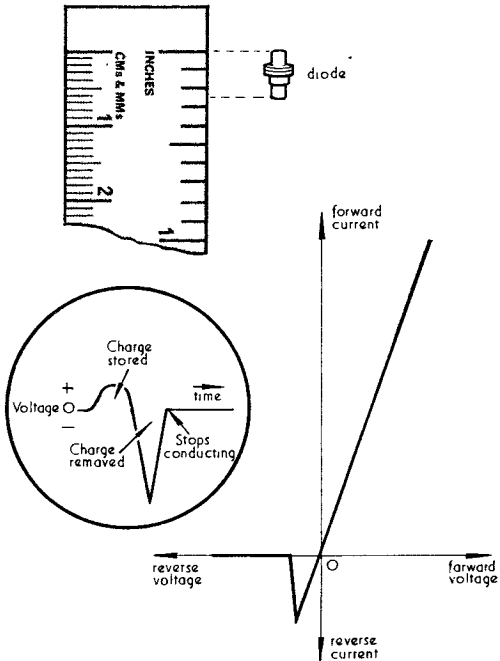


FIG 10. STEP-RECOVERY DIODE CHARACTERISTIC

have, in effect, an 'abrupt' junction between the p and n semiconductor materials. This construction ensures maximum switching speed in going into the reverse-biased condition. Therefore it returns most of the charge injected by the forward current. The net rectified current is very small and the fundamental-frequency power is converted into harmonic power with a high efficiency.

Whilst conducting in the forward direction, the step-recovery diode stores charge. When the applied voltage is reversed the diode conducts for a *brief period* in the reverse direction until the stored charge is removed and then *abruptly ceases conduction*. This is shown in Fig 10, which also indicates the origin of the name 'step-recovery'. This step-recovery transition is very rich in harmonics.

For frequency-multipliers of high output power, step-recovery diodes give the best efficiencies. Driven with a sinusoidal voltage the step-recovery diode produces harmonics whose amplitudes decrease as $\frac{1}{n}$, where n is the harmonic order. It will be remembered that the relationship for the varactor diode is $\frac{1}{n^2}$.

The step-recovery diode is most useful for high-order multiplication—the area where the varactor diode is limited. The big saving here is in the *number* of active devices needed. In most cases a transistor oscillator (or transistor-oscillator-multiplier) followed by a single step-recovery diode gives the required harmonic output. Many more varactors, or the inclusion of complex idler circuits, would be needed in the varactor-multiplier chain. Typical step-recovery diode performance figures are given in Table 3.

Frequency In	Multiplication	Frequency Out	Power In	Power Out	Efficiency
100MHz	x20	2GHz	50mW	15mW	30%
200MHz	x10	2GHz	15W	2W	12%
400MHz	x5	2GHz	15W	5W	30%
810MHz	x12	9.7GHz	400mW	10mW	2.5%
(from transistor doubler)	x10	8.1GHz	100mW	10mW	10%

TABLE 3. TYPICAL STEP-RECOVERY DIODE PERFORMANCE FIGURES

Tunnel Diodes

The tunnel diode, brief details of which are given in p342, may be used as an oscillator, an amplifier, or a mixer. The main research on tunnel diodes in recent years has concentrated on improvements in construction, cut-off frequency, and power-handling capacity.

A typical construction of a tunnel diode is shown in cross-section in Fig 11. The p-n junction is formed by alloying a small n-type tin-arsenic bead into the surface of a p-type germanium wafer. The wafer is soldered into the casing, which forms one contact, and the n-type bead is connected to the other end of the casing via a strip of silver foil. The silver foil is soldered to the bead to act as a *support* as well as a contact. The bead is very small (about 0.001 inch diameter) to ensure a low junction capacitance and a high cut-off frequency.

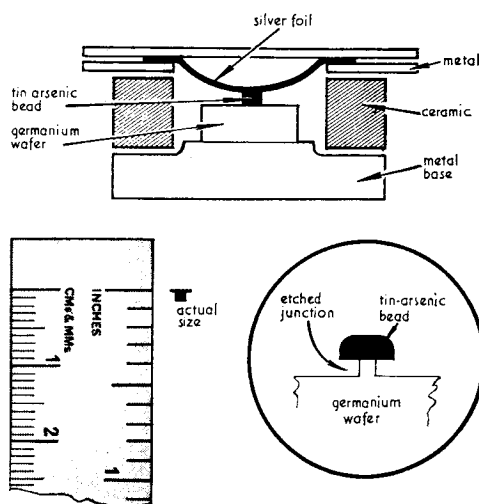


FIG 11. CONVENTIONAL TUNNEL DIODE CONSTRUCTION

Very often the junction capacitance is further reduced by *etching* round the junction as shown in the inset of Fig 11. Although this improves the cut-off frequency this process cannot be carried too far because the junction then becomes very fragile and 'top-heavy', with the appearance of a pebble perched on top of a pinhead.

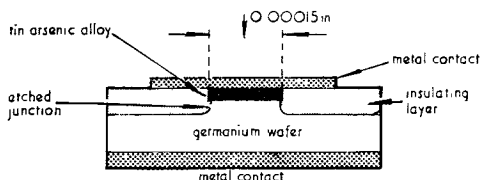


FIG 12. SOLID STRUCTURE TUNNEL DIODE

tances and high cut-off frequencies are possible with this solid structure tunnel diode.

Tunnel diode oscillator. Since a suitably-biased tunnel diode presents a negative resistance to its terminals at all frequencies up to cut-off, it is capable of resonating at the resonant frequency of any high-Q series-tuned circuit of smaller resistance connected across the terminals. The circuit shown in Fig 4, p343 is based on this principle.

It is also possible to construct a *resonant-cavity* tunnel diode oscillator capable of being tuned over a broad range of frequencies. The outline of a simple system is shown in Fig 13. The oscillation frequency is that of the cavity and may be varied by altering the setting of the tuning plunger; maximum oscillation frequency occurs with the plunger fully out.

Tunnel diode oscillators have been developed to operate satisfactorily at 100GHz. Currently-available tunnel diodes are usually quoted with cut-off frequencies in excess of 50GHz, at which frequency output powers of 200 μ W are possible.

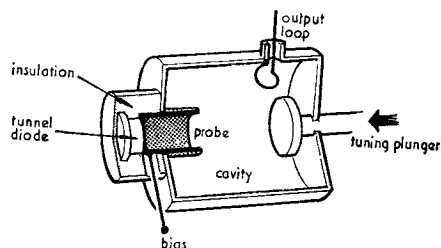


FIG 13. RESONANT-CAVITY TUNNEL DIODE TUNABLE OSCILLATOR

Tunnel diode amplifier. The tunnel diode is a "majority carrier" device and is not limited in frequency by the low diffusion velocity of minority carriers. Its high-frequency performance is limited only by its CR product and by the inductance of the package. By using small-area junctions, and semiconductor materials with high charge-carrier mobility, the CR product may be kept low to produce cut-off frequencies in excess of 50GHz. The useful upper frequency limit for tunnel diode amplifiers at the moment is of the order of 20GHz. Up to this frequency the tunnel diode amplifier can provide gains as high as 15db at noise figures as low as 3.5db.

Another advantage of the tunnel diode amplifier is its reliability, coupled with its small size and weight. By using strip-line components, very compact r.f. heads for X-band radar receivers have been developed. A typical system is shown in Fig 14.

Backward Diodes

Until fairly recently, metal-to-semiconductor point-contact crystal diodes were about the only devices used as mixers at microwave frequencies. However, mixing is possible with any device that has a non-linear current-voltage characteristic. Thus, with the introduction of tunnel diodes it became clear from the characteristic for this diode that mixing could be obtained by the

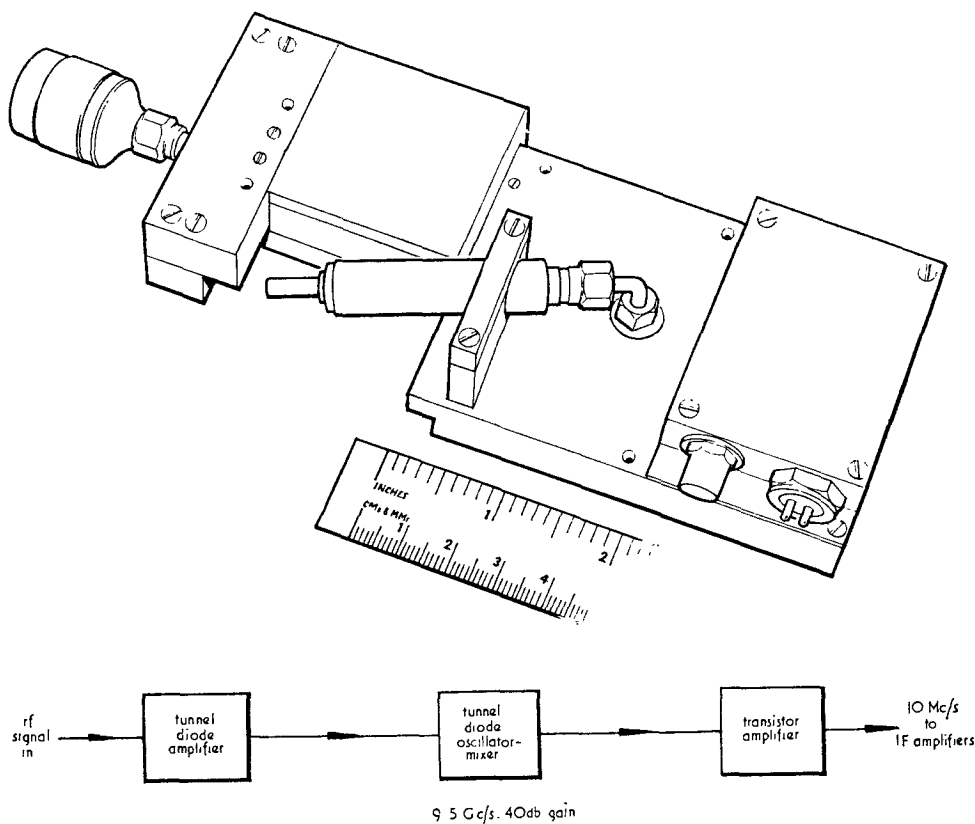


FIG 14. MICROWAVE RECEIVER INPUT STAGES USING TUNNEL DIODES

tunnelling effect of majority current carriers across the p-n junction. To operate satisfactorily as a microwave mixer the junction capacitance must be reduced as much as possible. This may be achieved by using a form of "point-contact" construction for the tunnel diode: a p-type gallium-plated gold wire may be *bonded* to an n-type germanium wafer as shown in Fig 15a overleaf. Note that this is an alloyed bond and *not* a *pressure contact* as in the conventional crystal mixer. It is also a semiconductor-to-semiconductor junction.

For mixing, we are not concerned with the negative-resistance property of the tunnel diode. This property is therefore reduced by suitable doping of the n-type germanium and by choosing a suitable value of current for the initial bonding of the cat's whisker.

A tunnel diode constructed as described is usually referred to as a *backward diode* because, unlike conventional diodes, the current when forward-biased is *less* than the reverse-biased current (Fig 15b). Another difference between the tunnel diode and the backward diode is that the tunnel diode is biased to operate in the negative resistance region of its characteristic whereas the backward diode operates at *zero bias*.

Fig 16 compares the characteristic of a backward diode used as a mixer with that of a conventional silicon point-contact crystal diode. For ease of comparison the backward diode characteristic has been reversed, the reverse current in a backward diode being equivalent to the forward current in a normal diode. Examination of these characteristics

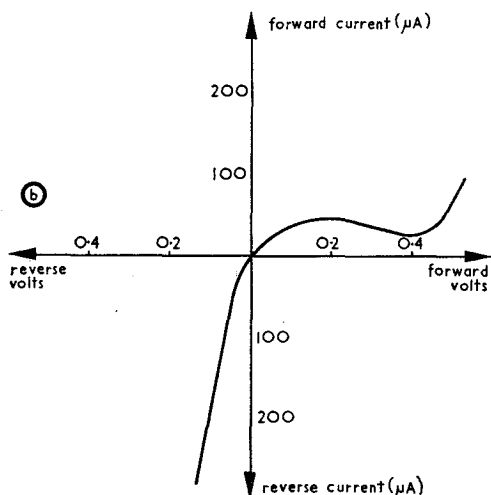
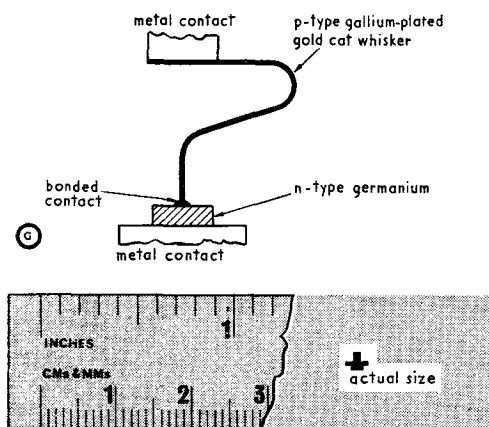


FIG 15. BACKWARD DIODE, TYPICAL CONSTRUCTION AND CHARACTERISTIC

Hot Carrier Diodes

With the increasing use of the microwave band there is a need for really fast, low-power switching devices capable of operating efficiently at the highest frequencies in use. The main limitation of p-n junction devices in high-frequency operation is associated with the minority carriers. In a varactor diode switch, for example, the minority carriers may be swept back across the barrier when the diode polarity reverses before they have time to recombine with holes or electrons (charge-storage effect). We have seen how the step-recovery diode tries to overcome this problem by building in a

shows that the backward diode has much greater curvature at zero bias and a steeper characteristic. This indicates the ability of the backward diode to rectify more efficiently than a silicon diode at small input power levels. Thus the backward diode gives greater conversion efficiency and has higher sensitivity.

However, the characteristics also show that the backward diode will tend to break down at lower levels of reverse voltage. The local oscillator power applied to a backward diode must therefore be *limited* to avoid causing excessive reverse current leakage. If the local oscillator power is correctly adjusted we then get an improvement in the noise figure of the backward diode (7db) as compared with the silicon diode (9db).

The other important factors for a mixer diode are the temperature range over which the diode operates satisfactorily, and the ability of the diode to withstand spike leakage from the TR cell. The backward diode has a slightly better temperature performance than the silicon diode, whilst in resistance to burnout the two devices are comparable. Thus, on balance, the backward diode offers significant advantages over conventional crystal mixers at microwave frequencies.

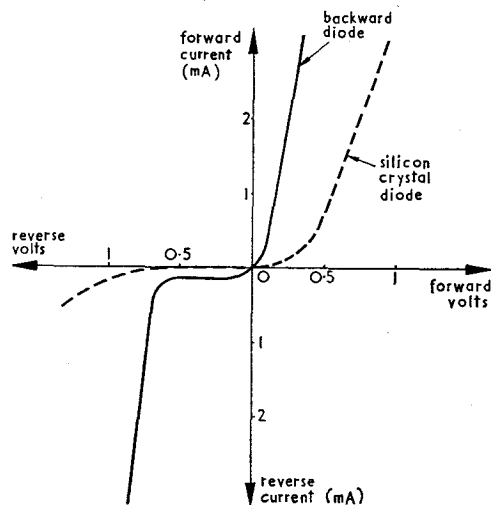


FIG 16. COMPARISON OF BACKWARD DIODE AND SILICON CRYSTAL MIXER DIODE CHARACTERISTICS

confining field to keep the charge carriers in the vicinity of the barrier. However, this is only a partial solution. Diodes have now been developed which reduce the charge-storage problem, simply by having no minority carriers. The 'hot carrier' diode is one such majority carrier device.

The hot carrier diode is little more than a modern form of metal-to-semiconductor diode and consists of a metal film (gold, silver, or platinum) deposited on a thin layer of semiconductor material, usually n-type silicon (Fig 17).

The 'energy level' in the semiconductor (see p355) is greater than that in the metal so that under forward-bias conditions electrons are *injected* from the n-type semiconductor over the potential barrier into the metal. These electrons are 'hot' in the sense that their energy *greatly exceeds* that of the electrons in the metal. Once into the metal the injected electrons quickly achieve equilibrium with the metal electrons.

In the reverse-bias condition the low-energy electrons from the metal cannot surmount the potential barrier. Thus, the hot carrier diode is not subject to charge-storage limitations. The

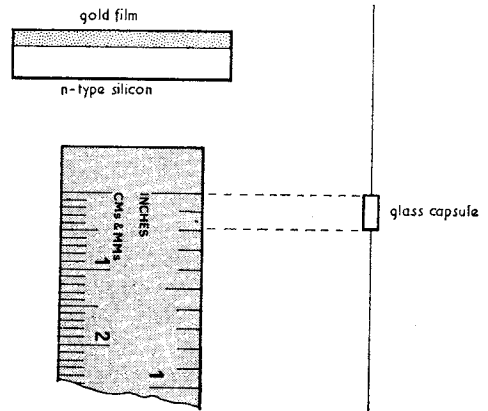


FIG 17. HOT CARRIER DIODE

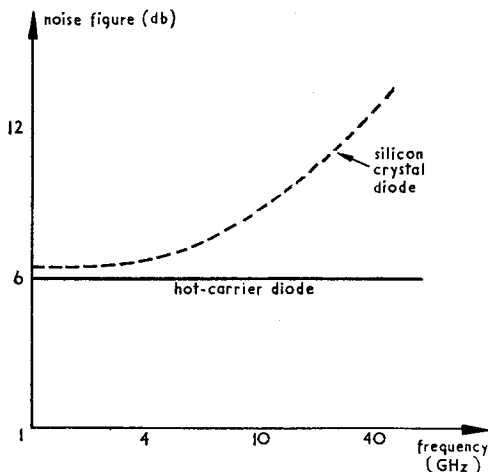


FIG 18. NOISE FIGURE VERSUS FREQUENCY, HOT CARRIER AND POINT-CONTACT MIXER DIODES

potential barrier between the metal and the semiconductor is often referred to as a 'Schottky barrier'. Because of this, hot carrier diodes are sometimes known as *Schottky-barrier diodes*.

With modern photo-chemical masking and diffusion techniques, very small devices with narrow junctions can be produced. Junction capacitances are usually less than 1 pF. Hot carrier diodes are available that are capable of passing up to 50mA forward current at 1V forward voltage, of withstanding up to 50V reverse voltage without breaking down, and of providing switching times of the order of 10 *pico*seconds.

The main use of the hot carrier diode is as a mixer at microwave frequencies. It has several advantages compared with the normal point-contact crystal diode mixer:

- Hot carrier diodes offer larger stable contact areas than do point-contact diodes and can, therefore, handle higher powers (typically 150mW c.w.).
- The reverse breakdown voltage is higher.
- The resistance to burnout is higher.
- There is a significant improvement in the hot carrier mixer system noise figure (Fig 18).

Hot carrier diodes are available for mixer applications at frequencies up to about 15GHz. Typical of these is a diode designed for use at frequencies around 8GHz. This diode mixer gives

a system noise figure of 7db when feeding into a 30MHz i.f. stage under a local oscillator power of 1mW at 8GHz. There is every hope that these figures will be improved as development proceeds.

PIN Diodes

PIN diodes are considered in p343 of this book. There we saw examples of their use as TR switches and as modulators. They may also be used as phase-shifters and attenuators. Typical p.i.n. diodes are shown in Fig 19. For most applications, the p.i.n. diode is connected *across* the waveguide, coaxial cable, or stripline to *control* the r.f. power being carried. The effect of the diode on the passage of r.f. energy depends upon the *bias conditions* of the diode. When it is reverse-biased it presents a high shunt impedance and has *little effect* on the r.f. energy being passed. When it is forward-biased it acts as a virtual short-circuit across the transmission line and the signal is *heavily attenuated*. The r.f. power is therefore controlled by altering the bias voltage applied to the diode.

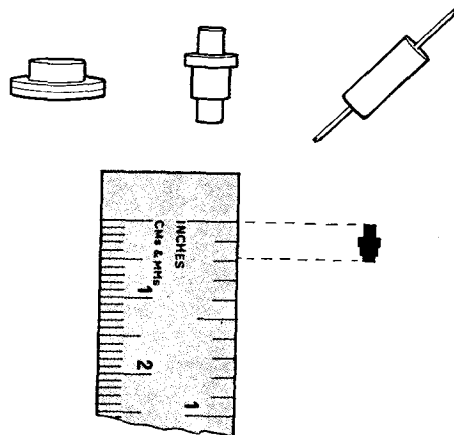


FIG 19. TYPICAL PIN DIODES

In practice the reverse-biased impedance is not infinite and the p.i.n. diode introduces some loss, known as *insertion loss*. Similarly, the forward-biased p.i.n. diode has a finite value of impedance which enables some r.f. energy to pass; in other words, the *isolation* is not perfect. The main research effort in p.i.n. diodes has been to reduce the insertion loss and improve the isolation, and to maintain these improvements at higher frequencies.

PIN diodes have been designed for use at frequencies up to 40GHz, with insertion losses as low as 0.1 db and isolation better than 30db. This means that the insertion loss of the diode is such that about 98% of the incident power passes to the output, and the isolation is such that only about 0.1% of the incident power reaches the output. For some applications, high peak power handling capability and high reverse breakdown voltage characteristics are required. PIN diodes capable of handling up to 10kW peak power with breakdown voltages of 1,500 volts are now available.

It is also possible to have devices using an 'array' of p.i.n. diodes suitably spaced along the length of the transmission line. This gives an improvement in the attenuation and control of the r.f. energy being passed. It also improves the *reliability* of the system because, in the event of the failure of one device, the reduction in the effectiveness of the system would be small. This built-in 'redundancy' is becoming common practice in many solid state equipments.

IMPATT Devices

The term 'IMPATT' (IMPact Avalanche and Transit Time device) has been used to describe a broad class of oscillating devices which use *impact ionization* to create an *avalanche* of charge carriers which, drifting in a *transit-time* region, produces a negative-resistance oscillation. Let us examine this rather complex statement in more detail.

IMPATT devices include those semiconductor diodes which exhibit negative resistance characteristics and which also depend for their operation on transit-time effects. Such diodes include the *silicon avalanche diode* and the *Read diode*. In general terms, an IMPATT device contains at least one semiconductor junction which is reverse-biased to such a value that the resulting electric field is sufficient to spontaneously generate electron-hole pairs by a form of *internal secondary emission* (avalanche or multiplication process).

Fig 20a shows the schematic outline of a reverse-biased Read 'diode' and Fig 20b illustrates a typical characteristic. The reverse bias creates a strong electric field across the transit-time region. The maximum field exists at the p-n junction and this is large enough to create avalanche conditions, producing a *pulse* of electron-hole pairs. The generated electrons are attracted by the applied field to the nearby n-region, whilst the positive charge carriers (holes) move through the transit-time region to the p-electrode. When the holes reach the p-electrode another avalanche pulse is generated, the *frequency* of the pulses depending upon the *transit time* of the charge carriers. The phase relationship between current and voltage is such that the device exhibits *negative resistance* characteristics. The diode can, therefore, be used as a negative-resistance oscillator.

The usable frequency range is partly determined by the transit time of the current carriers. Thus, for operation at, say, 10GHz the transit-time region would be very thin (about 0.000001 in). However, since the basic IMPATT device produces *pulses*, sinusoidal oscillations can be achieved only by mounting the diode in a microwave cavity. The same diode can be used with several different cavities to produce oscillations over a very wide frequency range. An experimental Read diode produced oscillations in the 2-4GHz band in a coaxial system, in the 7-12GHz band in X-band waveguide, and at 50GHz in millimetric waveguide.

IMPATT devices can provide the highest frequency or highest power output of any single solid state source at present available. A silicon avalanche diode has been developed to provide outputs of 1W c.w. at 12GHz with 8% efficiency. Outputs of a few milliwatts c.w. at 50GHz with 2% efficiency have also been obtained. There is every indication that these figures will improve to 20W in the range 5-20GHz, with efficiencies of 30%. The IMPATT device will then become a serious competitor to the travelling-wave tube.

Attempts have also been made to use an IMPATT device in an *amplifier* circuit. Gains of 20db, with 30MHz bandwidth, have been obtained at 10GHz. At present, however, the noise figure (50db) prohibits its use as an r.f. amplifier.

Gunn-effect Devices

When the voltage applied to a *thin slice* of gallium arsenide semiconductor exceeds a critical threshold value, periodic fluctuations (oscillations) of current result. For slices which are thin enough, the frequency of oscillation is in the microwave band. This effect is known as the 'Gunn effect' after its discoverer. The Gunn effect semiconductor is a 'bulk' device, so-called because there are *no junctions*.

Fig 21a shows a basic circuit of a gallium arsenide slice connected in series with a small resistor to a source of d.c. voltage. By measuring the current through the resistor it is found that the current is proportional to voltage up to a certain critical threshold value of voltage V_T . However, when the voltage is *increased above* V_T , the current *drops rapidly* from its value I_T

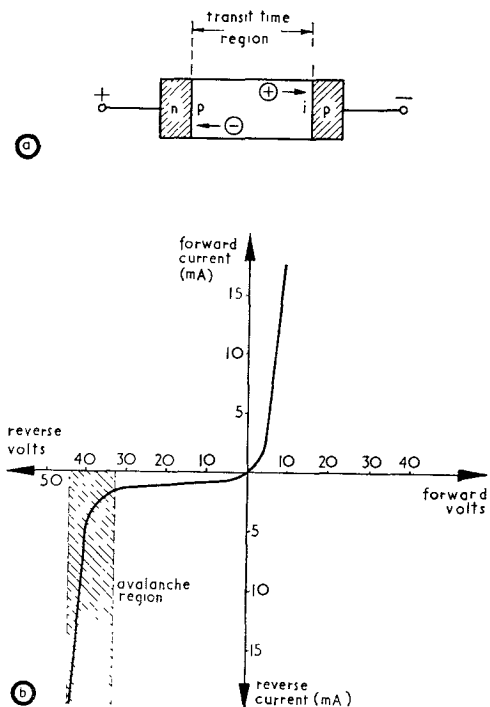


FIG 20. BASIC READ DIODE AND CHARACTERISTIC

to a "valley" current value I_v . The current then *remains* at the value I_v for a time t before rising to its original threshold value I_T . This is illustrated in Fig 21b.

So long as the voltage is *above* the threshold value V_T , Gunn-effect oscillations are generated and the current fluctuates *continuously* between the values I_T and I_v . Thus, oscillations are produced. The period of oscillation is related to the time t which, in turn, depends upon the *thickness* of the gallium arsenide slice; the oscillation period is completely independent of the applied voltage above V_T . For microwave operation, the semiconductor slices vary from 4 to 25 *millionths* of a metre thick (i.e. less than one-thousandth of an inch thick).

The cause of the current fluctuations is not yet fully understood. The generally accepted theory is that gallium arsenide is a semiconductor having *two energy bands* separated by a small energy gap (see p355). The lower energy band has as smaller mass associated with it and a *higher mobility* than the upper band.

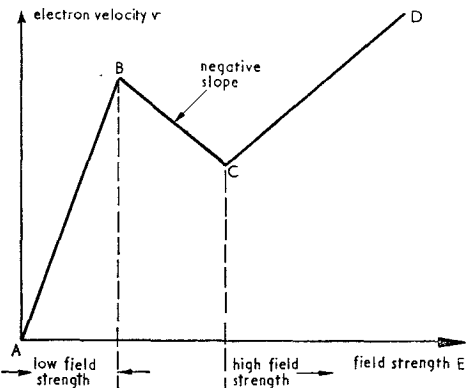


FIG 22. NEGATIVE-RESISTANCE COMPONENT OF GUNN-EFFECT DEVICES

In Fig 22 the electron velocity v in n-type gallium arsenide is plotted against electric field strength E , the latter being proportional to the *voltage* applied to the gallium arsenide slice. At *low* field strengths the electrons have low energy values and most of them are in the lower energy band, where the mobility is *high*. As the electric field is increased the electrons acquire more energy and become 'hot'. Thus, with the increase of field strength more electrons transfer to the upper energy band, where the mobility is

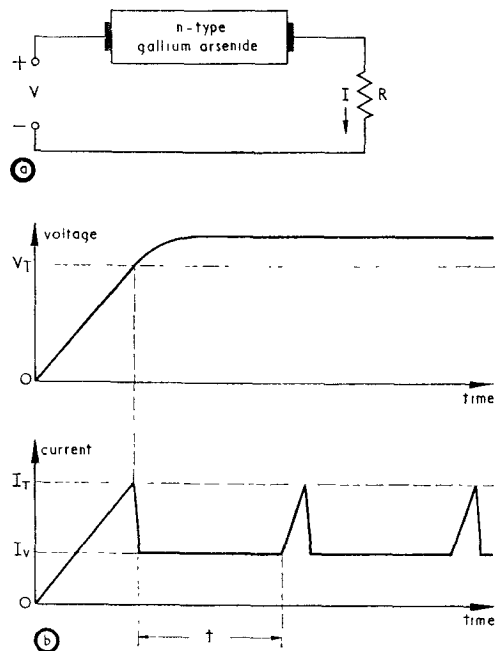


FIG 21. GUNN EFFECT IN GALLIUM ARSENIDE SEMICONDUCTOR

low. Fig 22 shows that at low field strengths the v - E curve has a steep slope AB (because of the higher electron mobility) and at high field strengths the slope CD has decreased. There must be a transition from slopes AB to CD. This is shown by the line BC which has a *negative slope*. The current fluctuations are connected with this region. Unlike the tunnel diode oscillator, however, the Gunn-effect device has no stable operating point along its negative slope.

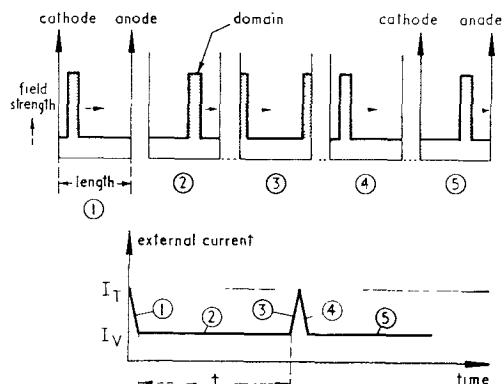


FIG 23. DOMAIN FORMATION AND MOVEMENT

If we were to examine the field distribution across the gallium arsenide slice at a given time after the threshold voltage V_T were exceeded, the result would be as shown in Fig 23. A very narrow region of high electric field (known as a *domain*) would be observed, moving along the length of the slice. During the time that the domain is *in transit*, the current in the external circuit remains at its valley level of I_V . It rises to the threshold value I_T as the domain moves into the 'anode' contact. A new domain is then formed at the 'cathode', and when it breaks away the external current returns once more to its valley level of I_V .

In the earliest Gunn-effect experiments the applied voltage was *pulsed* at a low p.r.f. to avoid overheating the semiconductor. Typical pulse durations were around $1\mu\text{s}$ and p.r.f.s about 100 p.p.s. Thus the output current consisted of a series of high-frequency pulses which were maintained for the duration of the input pulse (Fig 24).

With improved packaging techniques, heat dissipation has been improved so that *c.w.* operation is now possible. The available power output is, however, smaller with c.w. In pulsed operation, up to 200W peak power output can be obtained. For c.w. operation, power outputs are of the order of several milliwatts.

FIG 24. PULSED GUNN-EFFECT OSCILLATIONS

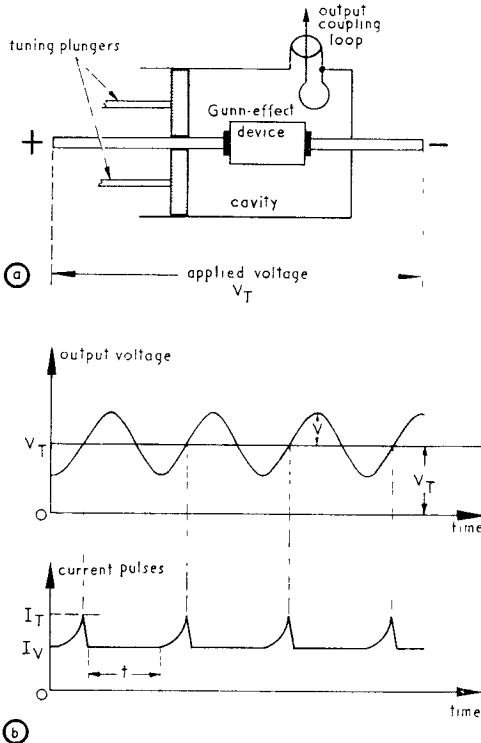
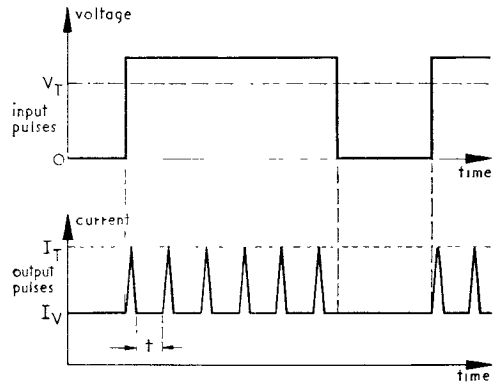


FIG 25. GUNN-EFFECT DEVICE IN RESONANT CAVITY

When the Gunn-effect device is connected in a basic circuit, the oscillation frequency is limited to the natural domain transit time. By placing the device in a *resonant cavity* it is possible to vary the oscillation frequency over a wide range.

When the device is placed in a resonant cavity (Fig 25a), the high Q of the cavity results in a *sinusoidal voltage* output even although the current through the device is non-sinusoidal and consists of a series of high-frequency pulses. Typical waveforms are shown in Fig 25b.

If the device is biased just to the threshold voltage V_T then the total voltage across it is the sum of V_T and the resonant a.c. component of amplitude V . A domain is launched just as the alternating voltage is passing through the V_T bias level in a *positive* direction. The current falls from its threshold level I_T to the valley level I_V and remains at this level for the domain transit time t . At this instant, as the domain enters the 'anode' the current rises again to I_T *only if the alternating voltage is at, or above, the V_T bias level*. If the period of oscillation of the cavity is greater than t , the above condition *does not apply* and the current rises only slowly to its critical value I_T , reaching this value when the alternating voltage reaches the V_T bias level.

The current then once more drops to the valley level I_v and the cycle begins again. Thus the frequency of oscillation is determined by the resonant frequency of the *cavity* and this may be in the range 1-40GHz. By making the cavity tunable the oscillation frequency may be varied over a given frequency range; typical tuning ratios are 1.5 : 1.

The factors to be considered when deciding on the suitability of any oscillator source include the operating frequency, the output power, and the efficiency. Table 4 lists these factors for available Gunn-effect oscillators.

Pulsed or CW	Frequency	Power Output	Efficiency
Pulsed	1GHz	200W peak	8%
Pulsed	3GHz	2.5W peak	7%
Pulsed	7.6GHz	1.5W peak	8%
CW	3GHz	60mW	6%
CW	5GHz	65mW	3%
CW	11GHz	110mW	3%
CW	35GHz	1mW	2%

TABLE 4. CHARACTERISTICS OF GUNN-EFFECT OSCILLATORS

Attempts are also being made to use the Gunn-effect in *amplifiers*. It may be seen from Fig 25b that by applying a bias voltage of a value *just below* the critical threshold value of voltage required for self-oscillation, a *single* domain can be launched by a small trigger pulse. This takes the voltage just over the threshold value. In this way an output pulse of *much larger* value can result and amplification has taken place. Amplifiers have been designed to operate in the range 2-12GHz with 5db gain at a noise figure of 20db. At the moment the high noise figure is limiting amplifier progress.

With research continuing into Gunn-effect devices it is clear that even higher frequency oscillations at higher power levels and greater efficiencies will result. The Gunn-effect device will then be in a position to offer a challenge to existing microwave power sources with additional advantages.

Microelectronics

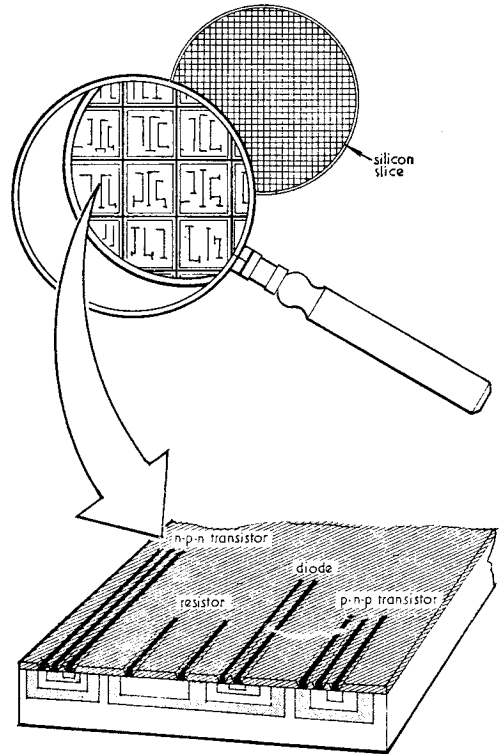
The fact that we can now make and interconnect several hundred transistors on a single silicon slice no bigger than 1 in diameter (see page 546) is of great importance. One question which immediately arises is, 'If this can be done with transistors, is it also possible to make and interconnect resistors, capacitors, diodes, and transistors on a single semiconductor slice—in fact, to make a *fully integrated circuit*?' The active research now being undertaken into the subject of *microelectronics* is attempting to provide an answer to this question.

We know that semiconductor solid state devices are much more reliable than their thermionic valve counterparts. In trying to improve reliability still further it is found, on examining solid state circuits, that one of the major causes of breakdown lies in the *interconnections* between components. This has always been one of the weakest links in any electronic circuit, and solid state equipments are no exception. However, if these separate interconnections could be eliminated such that they become an *integral part* of the highly-reliable semiconductor structure then the reliability of the circuit as a whole would be almost as high as that of the individual components. It is this which the *integrated circuit* part of microelectronics is trying to achieve.

Thus the main aim in microelectronics is to improve performance and reliability. Microelectronic techniques also produce extremely small and light circuits; but this is incidental and of secondary importance. The ultimate object is to provide trouble-free operation of an equipment.

Microelectronic circuits are currently available in three forms: integrated circuits, thin-film circuits, and hybrid circuits.

a. Integrated circuit. An integrated circuit (IC) is one in which a complete circuit is fabricated on a *single slice* of semiconductor, usually silicon. The circuit may include both *passive* components (resistors and capacitors) and *active* components (diodes and transistors). The production process is similar to that described on p.546 for the planar diffused transistor. A large number of individual steps, including successive diffusion and photo-chemical etching stages, is necessary to produce the integrated component structure shown in Fig 26. The single p-n junction may be used as a normal diode or it may be used as a capacitance (junction reverse-biased). The resistance value of the resistor depends upon the dimensions of the n-type diffusion layer and also upon its resistivity. Very high values of resistance can be obtained with this method. The body of the semiconductor slice may be used for lower values of fixed resistance. Microelectronic circuits fabricated in this way are often called '*monolithic*' devices. It is possible to construct many components on a single small slice of silicon, the interconnections being made by compression bonding to the alloyed contacts.



- — n-type silicon
- — p-type silicon
- ▨ — silicon oxide (insulator)
- — metal contacts

FIG 26. MONOLITHIC INTEGRATED CIRCUIT

b. Thin-film circuit. In thin-film microelectronic circuits various layers or films are deposited in succession on a glass substrate to form *passive* circuit components. Copper or gold metal films may be used to provide *conducting* surfaces for interconnecting wiring, inductors, and capacitor plates. Silicon monoxide can be deposited as a film to provide *insulation* or capacitor dielectrics.

Nickel-chromium alloys can be deposited to function as *resistors*. Fig 27 illustrates the basic processes in the production of a thin-film resistor. This basic procedure can be repeated to produce successive conducting, insulating, and resistive films of the required pattern. The complete 'component' may then contain several resistors, capacitors and inductors. The deposited films are extremely thin—of the order of one *millionth* of a metre. Thus, very compact circuit components result.

Thin-film passive components can normally be manufactured to more accurate values, with better tolerance, than their integrated monolithic circuit counterpart. Thin-

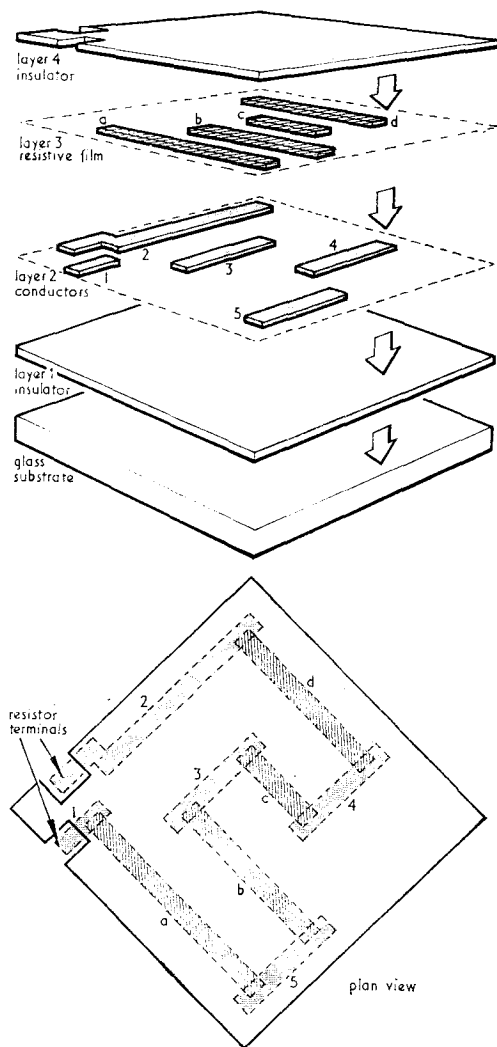


FIG 27. STAGES IN PRODUCTION OF A THIN-FILM RESISTOR

electronics has achieved considerable success, and many manufacturers have adopted it as the standard method for producing microelectronic circuits.

film resistance values can be higher than 1 megohm and can be produced with 3% tolerance. Capacitors of $0.1\mu\text{F}$ with 5 volts working voltage, can readily be produced. Furthermore, inductor windings have been successfully fabricated by thin-film techniques; in the monolithic integrated circuit there is no easy way of providing inductance.

The *active* components used with thin-film microelectronic circuits are usually *unpackaged* semiconductor diodes and transistors, referred to as “chip” devices. The use of chip devices has been made possible by applying planar diffusion techniques to the semiconductor. The final insulating diffusion effectively “seals” the device. The chip devices are attached to the thin-film wafers by compression bonding. A typical thin-film circuit connected to chip diodes and transistors is shown in Fig 28. Note the size compared with a pin.

Research is now going on to develop thin-film *active* devices. Some success has been achieved in this although a satisfactory manufacturing process has not yet been fully developed. If successful thin-film diodes and transistors can be made available to replace existing chip devices, the thin-film circuit will also be a truly “integrated” circuit. It will then be fully competitive with monolithic devices.

c. Hybrid circuits. Hybrid microelectronic circuits have also been produced, combining the relative advantages of monolithic and thin-film circuits. A hybrid circuit consists of passive thin-film elements deposited on the insulating region of a diffused silicon slice, the latter containing the active components in monolithic form. This approach to micro-

Microelectronic techniques have been applied successfully to the *microwave* region. A four-stage microwave amplifier using thin-film techniques and chip devices has been produced; this operates satisfactorily up to 4GHz giving a gain of 40db with a noise figure of 3.5db. It has a performance equal to that of a low-noise travelling-wave tube and offers longer life and greater reliability. Radar receiver r.f. heads, designed for operating at X-band, have also been pro-

duced successfully by microelectronic techniques. One r.f. head of 7 cubic inches volume and weighing only 10 ounces has been developed to handle 5kW peak power at a bandwidth of 200MHz, and with a noise figure of 9db.

Fig 29 illustrates a good example of the application of microelectronics to microwave equipment. The equipment shown is the transmitter-receiver and decoder of a micro-miniature IFF Mk 10 ground installation. The integrated circuit construction ensures high reliability coupled with low-volume packaging.

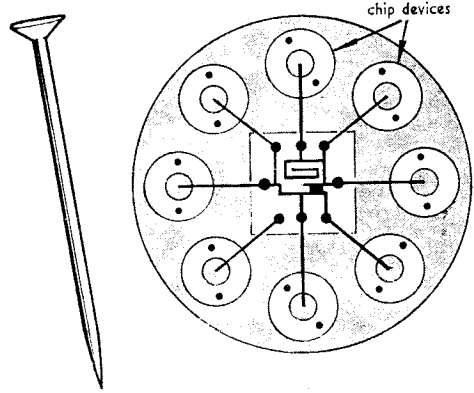


FIG 28. TYPICAL THIN-FILM CIRCUIT CONNECTED TO CHIP DEVICES

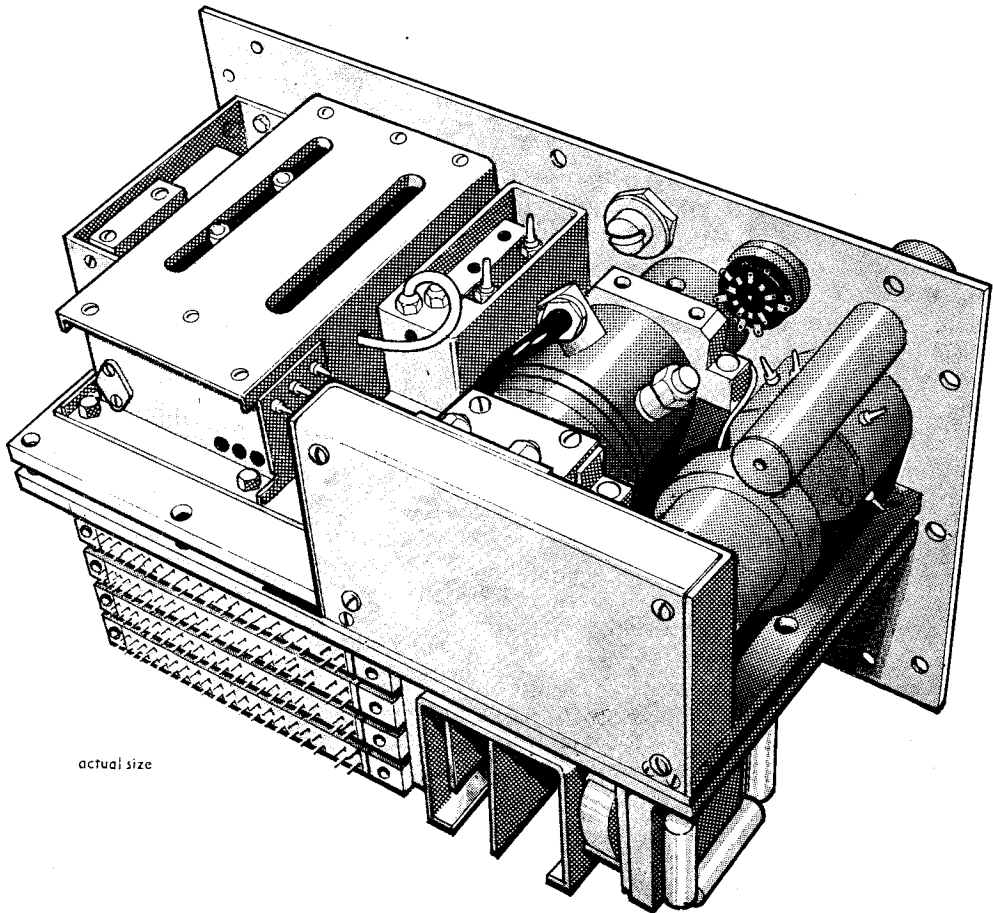


FIG 29. EXAMPLE OF SOLID STATE INTEGRATED CIRCUIT EQUIPMENT

DOFIC

As a result of research work now being undertaken in this country, an entirely new design concept in integrated circuits has emerged which may well ultimately replace existing monolithic integrated circuits. The new device, which is known as a DOFIC (Domain Originated Functional Integrated Circuit), makes use of basic *bulk effects* in semiconductor material, similar to the Gunn effect in gallium arsenide.

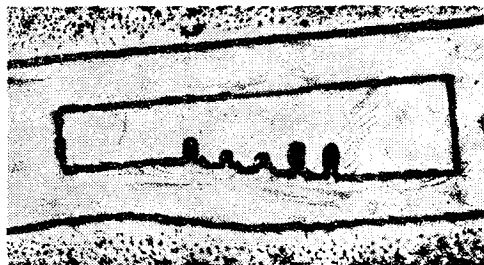


FIG 30

DOFIC ANALOGUE-TO-DIGITAL CONVERTER

the DOFIC, complex electronic functions can be derived from a *single* bulk effect device. The possibility therefore arises of replacing circuits containing many interconnected electronic components by a single piece of semiconductor material a few hundredths of an inch long.

Just as in Gunn-effect devices, the DOFIC is able to produce stable high electric field "domains" moving over a distance that is long compared with the width of the domain. These domains are launched when an applied voltage, exceeding a certain threshold value, is applied across the ends of the device (see p.560 Fig 23). The current through such a device will change as a domain encounters a change in conductivity (caused by variations in doping) or changes in cross-sectional area. Thus, by defining the conduction path in these ways, an output current waveform of almost any shape may be produced. Such conductivity profile shaping determines the *static characteristic* of the device. A basic DOFIC analogue-to-digital converter, as it would appear under a microscope, is illustrated in Fig 30. The "digital" profile (the five indentations) and the overall slope combine to give the required current output waveform.

Perhaps of even greater significance is the fact that *dynamic control* of the device is possible (i.e. whilst a domain is *in motion*). The dynamic control may be obtained by varying the instantaneous bias applied to the device. Thus, a domain can be arrested at any point along its drift path. In addition, the point along the path at which the domain is removed can be made proportional to the applied bias. In the simple device shown in Fig 30 the number of output current pulses is proportional to the applied bias (analogue-to-digital conversion). This is illustrated in the graph of Fig 31. The time scale of this graph indicates operation in the microwave region.

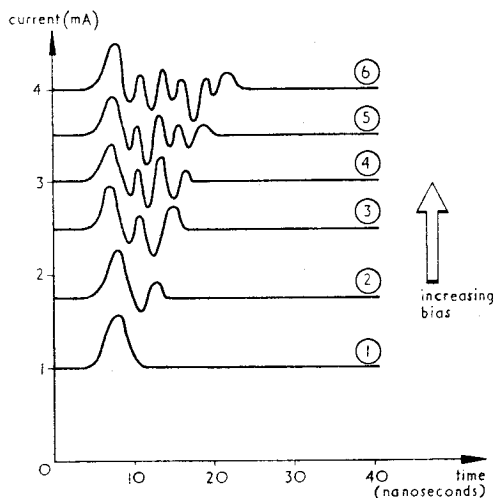


FIG 31.
CONTROL OF DOFIC BY VARIATION OF BIAS

So far we have been concerned with the *total* current through the device. However, it is also

possible to detect the high fields caused by the domain by placing probes close to the surface of the semiconductor. This probe may itself have an additional DOFIC profile built into it. The sweeping domain in the main DOFIC will then induce voltages as it moves past the probe, the latter being shaped to produce any desired time spacing of the output pulses.

A further significant characteristic of the device is that the domain can be made to excite light as it passes through the semiconductor. This makes it possible to obtain optical read-out of the waveform by detecting the generated light. This possibility brings prospects of entirely new types of display devices which could have far-reaching effects in radar display systems.

We have seen that the DOFIC can be used to produce electronic outputs which, up till now, have only been obtainable by using relatively complicated arrangements of discrete or integrated circuit components. As yet the research is in an early stage, but already the prospects are such that we may be witnessing the birth of the next generation of electronic devices.

Summary

In this chapter we have dealt very briefly with available microwave semiconductor devices. Ways in which progress has been made have been indicated and prospects for the future have been examined. However, research and development are proceeding at such a pace that it is probable that even before this chapter is in print some new advance in the technology will have been announced. Because of this, it is difficult to guess which devices will prosper and which will die. For example, it is possible that advances in f.e.t. design will be such that the f.e.t. will supersede the conventional bipolar transistor for many applications. Time will tell.

In the meantime, sufficient information has been given in this chapter to indicate the probable trends. Only the *fringe* of each subject has been examined, however, and more detailed study of a particular device will be required as and when it appears in an equipment. When working on such an equipment always refer to the appropriate equipment AP.

DEVELOPMENTS IN MICROWAVE VALVES, FERRITE DEVICES, AND DELAY LINES

Introduction

Although the development of microwave semiconductor devices (Chap. 1) has had great emphasis placed upon it, there are other fields in which research and development have been just as great, if not as spectacular.

Considerable advances have been made in recent years in the development of many microwave components. Microwave valves have been improved in terms of available power output, efficiency, bandwidth, and gain, at higher frequencies. Ferrite devices, including circulators, isolators, and phase-shifters, have been improved to such an extent that their field of operation is now in the megawatt peak power region. Rapid advances have also been made in the development of microwave delay lines. This chapter deals with some of the developments in these fields.

MICROWAVE VALVES

Types and Applications

At the moment, for *high-power* microwave sources (greater than about 10W c.w.), there is no alternative to microwave valves; no solid state device can as yet supply this power. On the other hand, in the *low-power* region (less than about 10W c.w.), the impression is sometimes given that valves have been completely ousted by solid-state devices. Whilst it is true that there is a trend towards this position, it will be many years before low-power valves are completely superseded by solid-state devices in the microwave region.

Fig. 1 shows the power/frequency areas where valves and solid state devices may be expected to be used. There is much overlapping, so that there is no definite dividing line where one can say, "Valves will be used here, and solid state devices there". Future equipments will use both microwave valves and solid state devices.

The four main types of microwave valves used as *amplifiers* are planar triodes, travelling-wave tubes, klystrons, and crossed-field amplifiers (e.g. the amplatron). For *generation* of microwaves, we have magnetron oscillators, reflex klystrons, and backward-wave oscillators. All these valves have been considered earlier in this book. What we now wish to do is to examine those fields in which improvements have been made and to note prospects for future development.

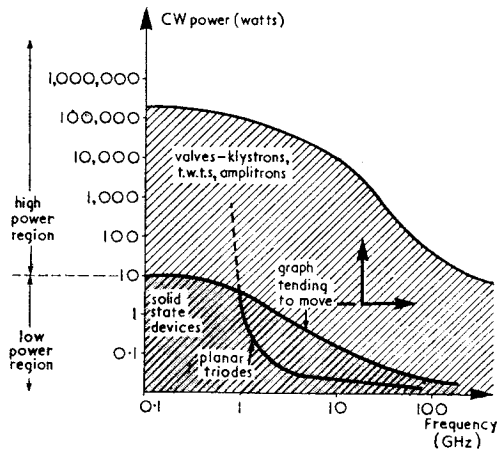


FIG. 1. USEFUL POWER/FREQUENCY AREAS FOR VALVES AND SOLID-STATE DEVICES

Microwave Planar Triodes

An introduction to planar triodes is given on p231 of this book. It is stated there that planar valves have their greatest usefulness at frequencies below 1GHz, and that at these frequencies very high values of average power can be produced. Recent developments have resulted in satisfactory operation of planar valves at very much higher frequencies, up to 20GHz.

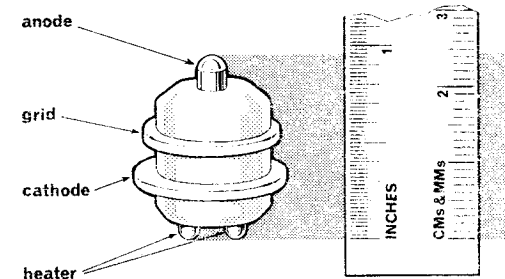
One of the main factors limiting the operation of planar valves at the higher frequencies is the loss associated with transit time effects. To reduce this loss, the grid-cathode spacing must be very small, and in modern planar triodes the spacing has been successfully reduced to 0.0005 in. (0.0125 mm.).

For efficient operation at the higher frequencies the cathode current density must also be high, of the order of several amperes per square centimetre (A/cm.²). Power output and efficiency are both improved with a high cathode current density (Fig. 2).

Cathode temperatures must be high to give the required current density, and the materials used for the cathode must be able to withstand these high temperatures. A new form of cathode construction, known as the bonded-heater cathode structure, has enabled these conditions to be met, whilst maintaining a reasonable "life" for these valves of about 5,000 hours.

A typical planar triode is illustrated in Fig. 3. When this triode was used with resonant cavities tuned to 1.2, 2.5, and 4.0GHz, measurements showed that the efficiency *decreases* and anode current *increases* with increasing frequency. The valve was used under pulsed conditions, the duty factor being as shown in Fig. 3.

The advantages of high current densities become even more apparent in valves using water-cooled anodes. A typical water-cooled ceramic triode for use at 1.3GHz provides 1kW c.w. power output at 60% efficiency with a gain of 13db; the valve is 1³/₈ in. long by 1 in. diameter (3.5 cm. by 2.5 cm.).



f (GHz)	Va (volts)	Ia (amps)	peak power (kW)	efficiency (%)	duty factor
1.2	1500	1.5	1.0	45	0.005
2.5	1500	1.8	1.0	33	0.004
4.0	1500	2.0	0.8	26	0.003

FIG. 3. TYPICAL PLANAR TRIODE AND CHARACTERISTICS

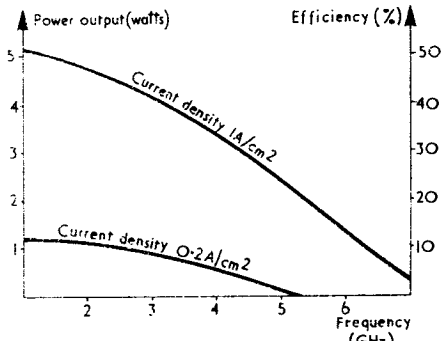


FIG. 2. POWER AND EFFICIENCY PLOTTED AGAINST FREQUENCY FOR A TYPICAL PLANAR TRIODE

Planar valves have been designed as local oscillators at X-band (10GHz) and at K-band (20GHz), where power outputs of 50mW c.w. at efficiencies of 20% have been obtained.

One new technique permits the operation of these valves in a dual transmitter/local oscillator role. For example, at X-band a planar triode has been developed to deliver a 1μs 50W transmitter pulse followed by a 50μs 50mW local oscillator pulse at a duty cycle of 0.25. The local oscillator frequency is displaced from that transmitted by an amount equal to the receiver i.f. The voltage applied to the valve anode is switched rapidly between high and low values to obtain the required outputs.

To summarize, we can say that metal-ceramic planar valves have now been developed to provide satisfactory operation at all frequencies up to about 20GHz. Power output, gain, and efficiency all decrease as the operating frequency increases; c.w. power outputs fall from tens of

kilowatts at 400MHz to a few milliwatts at 20GHz; gains fall from 40db to 5db; and efficiencies from 70% to 10%. Nevertheless, these valves continue to be used for many modern applications.

Travelling-wave Tubes

Brief details of the travelling-wave tube (t.w.t.) are given on p263. It was shown there that the t.w.t. is characterized mainly by its *broad bandwidth* and its *high gain*. Bandwidths of 2 : 1 are common (e.g. operating over 2-4GHz, or 5-10GHz), and some t.w.t.s are available with operating frequency ranges of 4 : 1. This feature, coupled with its high gain, reliability, and ability to operate at high peak power levels with a minimum of volume and weight, has firmly established the t.w.t. amplifier in radar systems, in e.c.m. systems, and in wideband data transmission and communication systems.

At the moment, however, the t.w.t. is inferior to other microwave amplifiers in certain aspects. Its *efficiency* is much lower than that of the crossed-field amplifiers (e.g. amplitrons), and the klystron can handle *higher average powers* than the t.w.t.

In most applications where microwave amplifiers are needed, maximum efficiency is required. This is particularly true for systems where power supply and weight requirements have to be limited. Since overall efficiency is defined as r.f. output power divided by total d.c. input power, an inefficient system will require a larger and heavier power supply than an efficient system.

The total d.c. power input includes the power supplied to the collector or anode, the heater power, and the power needed to make up the losses in the circuit. The latter two form a much larger fraction of the total power at low power levels than they do at high power levels. Thus, under this definition of efficiency, a high power amplifier will tend to provide a higher efficiency figure than a low power amplifier.

Until recently, the efficiencies of low- and medium-power t.w.t.s were of the order of 25%. By careful design of the valve structure, losses have been significantly reduced, and efficiencies greater than 40% with an output power of a few watts are now possible. The improved design has brought other benefits. Some t.w.t.s are now able to operate satisfactorily with low cathode temperatures. This has improved t.w.t. reliability and has resulted in a long life of more than 50,000 hours for such valves. This factor is of great importance for such applications as satellite transmitters and unmanned microwave repeaters.

TWTs can be manufactured as low noise amplifiers for use in r.f. stages of receivers. They are also available as low-, medium-, and high-power amplifiers for transmitters (Table 1).

TABLE 1. COMPARISON OF POWER TWTs

	Low power	Medium power	High power
Application	Airborne radar	Driver for klystron in large ground radar	Output stage of large ground radar
Power	20W c.w.	5kW peak	2MW peak
Type	Helix	Modified helix	Coupled cavity
Bandwidth	5-11GHz	11-14GHz	2.6-3GHz
Beam Voltage	3.3kV	10kV	120kV
Beam Current	0.08A	2A	75A
Gain	40db	50db	35db
Size	9 in. $\times$ 1½ in. (22.85cm. $\times$ 3.8cm.)	24 in. $\times$ 3½ in. (61cm. $\times$ 9cm.)	42 in. $\times$ 8 in. (1.07m. $\times$ 20cm.)
Weight	1½ lb. (0.57kg.)	35 lb. (15.9kg.)	100 lb. (45.36kg.)

In general, t.w.t.s are built round one of two basic types of interacting circuits: the helix, and the coupled cavity. Helix tubes can give wide bandwidths because of their non-resonant nature, but stability becomes a problem at peak powers greater than about 1kW. Between peak powers of 1kW and 100kW, a 'ring-bar' type of helix is used. Above 100kW peak power, or where the *average power* is high, coupled cavity circuits are used because of the ease of cooling such systems. Since these cavities are resonant, t.w.t.s using this type of circuit have a smaller bandwidth than helix types, of about 15% maximum. Even with this reduced bandwidth, cavity-coupled t.w.t.s can rival klystrons at peak power levels greater than 100kW. At low

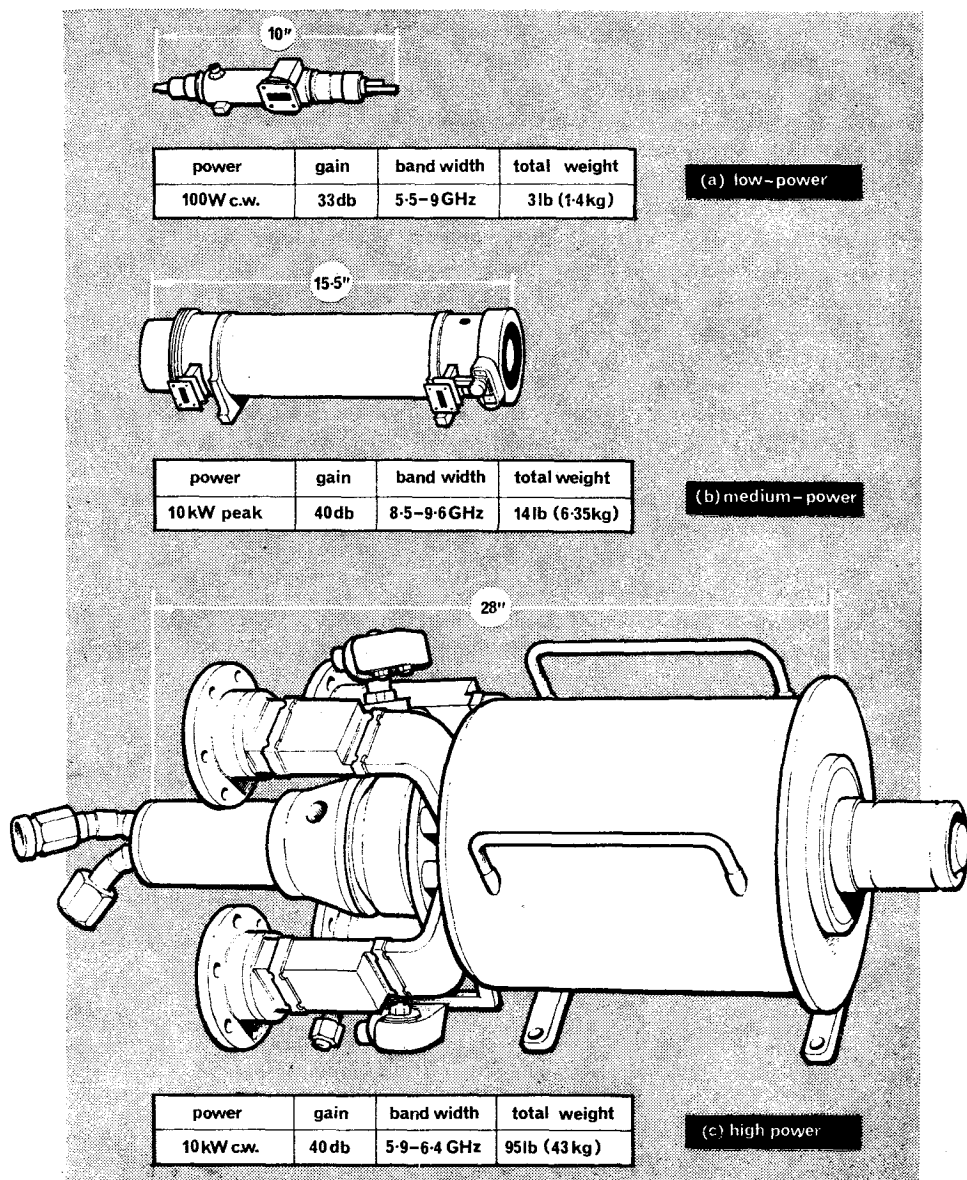


FIG. 4. TYPICAL POWER TWTs

power levels, permanent magnet focusing is used. High-power t.w.t.s usually require a solenoid for adequate focusing power (Fig. 4).

Low noise t.w.t.s for use in receivers have been successfully developed for operation at frequencies greater than 100GHz. Noise factors of available t.w.t.s range from 3db at 1GHz to 10db at 26GHz. The lowest noise factor is obtained at the lower power levels, corresponding to low beam current. A 50mW amplifier will have a lower noise factor than a 10W amplifier. Fig. 5 illustrates how the noise factor of 50mW t.w.t.s tends to vary with frequency.

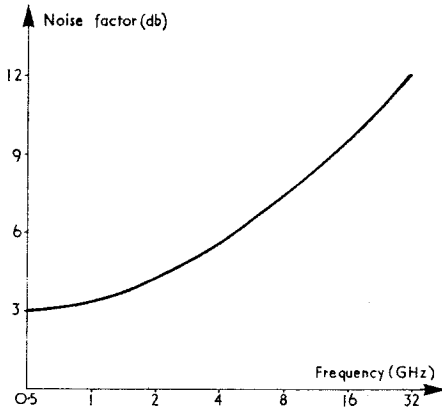


FIG. 5. VARIATION OF NOISE FACTOR WITH FREQUENCY, TYPICAL LOW-POWER TWTs

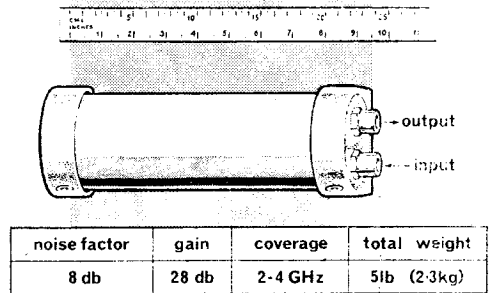


FIG. 6. BATTERY-OPERATED TWT

Fig. 6 illustrates a *battery-operated* t.w.t. for use as a low-noise r.f. amplifier in receivers. The battery which supplies an input power of 650mW at 28V d.c. gives 11 hours continuous operation. The low input power requirement is made possible by using a new cathode design that uses one-tenth of the heater power of a conventional t.w.t.

Klystron Amplifiers

Brief details of the operation of multi-cavity klystron amplifiers are given on p257. Just as the outstanding feature of the t.w.t. is its broadband capability, so that of the klystron is its ability to handle *high average levels of power*. In this the klystron is rivalled only by grid-controlled planar valves and then only at frequencies below 1GHz. CW powers of 175kW at 3GHz have been achieved in a klystron, and even this high level does not indicate the maximum possible: a multi-cavity klystron is being developed to produce 1MW c.w. at 8GHz.

The two main developments in klystron amplifiers have been the introduction of *electrostatic focusing* to reduce weight and size, and the application of '*extended-interaction cavities*' to improve efficiency and bandwidth.

Conventional klystron amplifiers employ a narrow beam which is made to pass through a series of cavity resonators with as little interception as possible. Some means of focusing is necessary to prevent excessive spreading of the beam by mutual repulsion of electrons. Magnetic focusing is common, using either a permanent magnet or a solenoid. In either case, the focusing system adds considerable bulk and weight to the valve. Furthermore, a solenoid requires an extra power supply which increases still more the weight and size of the system. Such factors are important in mobile, airborne and space applications.

It is, of course, possible to focus a beam by other means, as in the electrostatic c.r.t. The recent application of *electrostatic focusing* to klystron amplifiers gives considerable savings in size and weight of the system. The principle of operation of the electrostatically-focused klystron is illustrated in Fig. 7, overleaf. Ring-shaped lens electrodes are inserted between the cavities,

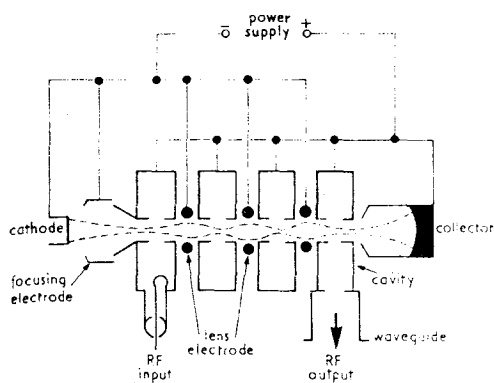


FIG. 7. ELECTROSTATIC FOCUSING IN KLYSTRON AMPLIFIER

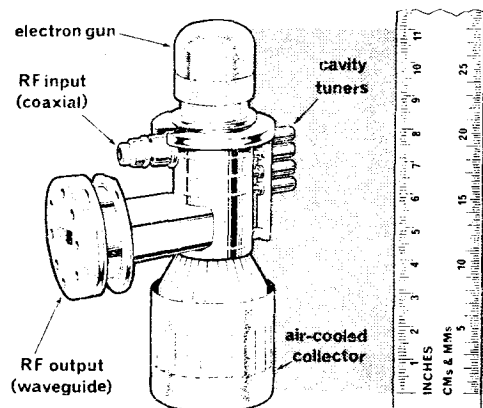


FIG. 8. ELECTROSTATICALLY-FOCUSED KLYSTRON

and the voltages applied to the various electrodes are adjusted to provide the required focusing. In practice, the lens electrodes are usually at the same voltage as the cathode, whilst the cavities are at the same voltage as the collector.

A typical electrostatically-focused klystron is illustrated in Fig. 8. It is a four-cavity klystron which is mechanically-tunable over the range 4.4–5GHz. It operates at a cathode voltage of -8 kV and a beam current of 0.5 A . The valve provides a minimum power output of 1 kW c.w. and a gain of 42 dB for a 3 dB bandwidth of 6 MHz . Its dimensions are as shown in Fig. 8 and it weighs only 17 lb. (7.7 kg.). A comparable conventional klystron with magnetic focusing would be much larger and could weigh as much as 65 lb. (29.5 kg.).

In the conventional multi-cavity klystron the bunching of the electrons at the cavity lips results in energy being extracted from the electron beam. In the *extended-interaction cavity klystron* an attempt is made to apply the extended interaction technique of the t.w.t. whereby the interaction between the beam and the cavities is extended over a *longer time*. In this way more r.f. energy is extracted from the beam and the *efficiency* of the klystron increases. The extended-interaction cavity incorporates a *slow-wave structure*, as in the backward-wave oscillator (p266), and this effectively reduces the velocity of the r.f. energy to give a longer interaction time between the beam and the wave. High conversion efficiencies of up to 65% are possible using this technique, and at the same time the high gain, stability, and long life of the klystron are retained. The extended-interaction output cavity also increases the power-handling capabilities of the klystron and, because of the slow-wave structure, the bandwidth of the device is improved.

The demonstration of higher efficiency and larger bandwidth by this technique has led to the development of *hybrid* klystron/t.w.t. amplifiers. One such valve, with a klystron driver section and an extended-interaction, t.w.t. type output section, has been designed for operation at $2.78\text{--}3.22\text{ GHz}$. It produces a peak power output of the order of 10 MW at 25 kW mean power with a gain of 38 dB .

Multi-cavity klystron amplifiers are available at frequencies from about 400 MHz to 200 GHz . Peak powers produced range from a few kW to about 40 MW ; average powers range from a few mW to about 175 kW . The gain is typically 40 dB and efficiencies can be as high as 65% . Klystrons have a long operational life, in excess of $10,000$ hours. They are ideally suited to those applications where high average powers, high gain, and good stability are required. The size reduction achieved by electrostatic focusing, and the improvement in efficiency and bandwidth achieved by the use of extended-interaction cavities have further increased the usefulness of these valves.

Crossed-field Amplifiers

The adjective '*crossed-field*' is a term used to describe those devices which depend for their action on the fact that they have two fields, an electrostatic (E) field, and a magnetic (H) field, whose axes are *at right angles* to each other. Thus, the magnetron is a crossed-field oscillator; the amplatron (p275) is a crossed-field amplifier.

We saw on p266 that as an electron beam passes the gaps in a slow-wave structure, oscillatory fields are set up across the gaps to produce both backward and forward waves. In the amplatron crossed-field amplifier it is the *backward wave* which is used, and the forward wave is deliberately absorbed in a non-reflecting termination so that it plays no further part in the operation. However, as in the t.w.t., the valve can be constructed such that it is the *forward wave* which is amplified and the backward wave which is absorbed. Thus we have both backward-wave and forward-wave crossed-field amplifiers.

Both types are characterised by *high efficiency*, often of the order of 85%. The crossed-field amplifier can deliver r.f. power at higher efficiency than the klystron or t.w.t. because the r.f. fields keep the electrons bunched even as energy is being extracted from the device. It also exhibits very stable *phase* characteristics, a factor of great importance in coherent radar systems (see p 502).

However, its average power-handling capacity is less than that of the klystron, its bandwidth of about 10% of the operating frequency is much less than that of the t.w.t., and its gain is less than either.

Recent efforts in crossed-field amplifier design have concentrated on improving those factors which compare unfavourably with other microwave amplifiers. The gain of such amplifiers has been improved in recent years from less than 10db to about 20db. This figure is still lower than that of comparable klystrons and t.w.t.s (typically greater than 40db) but there is hope of further improvement. Bandwidth is still limited, but average power handling capacity has been improved. Size and weight reductions have also been obtained: a 25W amplatron has been developed for space use weighing only 2 lb. (1kg.) including magnet.

Because of back-heating of the cathode by electron bombardment it is usually necessary to reduce or switch off the heater current of crossed-field amplifiers once the emission temperature has been reached. In fact, many modern crossed-field amplifiers operate with a *cold cathode*.

Crossed-field amplifiers are available at frequencies ranging from 500MHz to 200GHz, at average power levels ranging from a few watts to about 100kW, and at peak power levels up to 25MW. Details of typical amplifiers are given in Table 2. Fig. 9 overleaf illustrates three of the amplifiers quoted in this table.

Frequency	Peak power	Average power	Efficiency	Gain	Other remarks
a. L-band 1.2GHz	100kW	3kW	45%	15db	Forward-wave type; weight 35 lb. (16kg.) (see Fig. 9).
b. S-band 3GHz	120kW	4kW	60%	17db	Designed for use in sideways-looking radars; weight 50 lb (22.7kg.)
c. C-band 5GHz	3MW	10kW	65%	7db	Backward-wave type; weight 60 lb. (27kg.).
d. X-band 10GHz	500kW	700W	50%	14db	Weight 35 lb. (14kg.) (see Fig. 9).
e. Ku-band 16GHz	100kW	3kW	35%	18db	Weight 10 lb. (4.5kg.) (see Fig. 9).

TABLE 2. DETAILS OF CURRENT CROSSED-FIELD AMPLIFIERS

To complete this brief review of developments in microwave amplifier valves it may be of interest to compare some of their main characteristics. This is done in Table 3.

Valve	Range of Frequencies	Peak power (max.)	Average power (max.)	Gain (typical)	Efficiency (typical)	Remarks
Planar triode	300MHz to 20GHz	—	500kW to 1kW	15db	25%	Efficiency and power output decrease rapidly with increase in frequency. Mainly useful below 2GHz.
TWT	500MHz to 100GHz	10MW	30kW	40db	40%	Very broad bandwidth. Low average power.
Klystron	400MHz to 200GHz	40MW	>175kW	40db	60%	High average power. Narrow bandwidth.
Crossed-field amplifier.	500MHz to 20GHz	25MW	100kW	15db	80%	High efficiency. Low gain.

TABLE 3. COMPARISON OF MICROWAVE AMPLIFIER VALVES

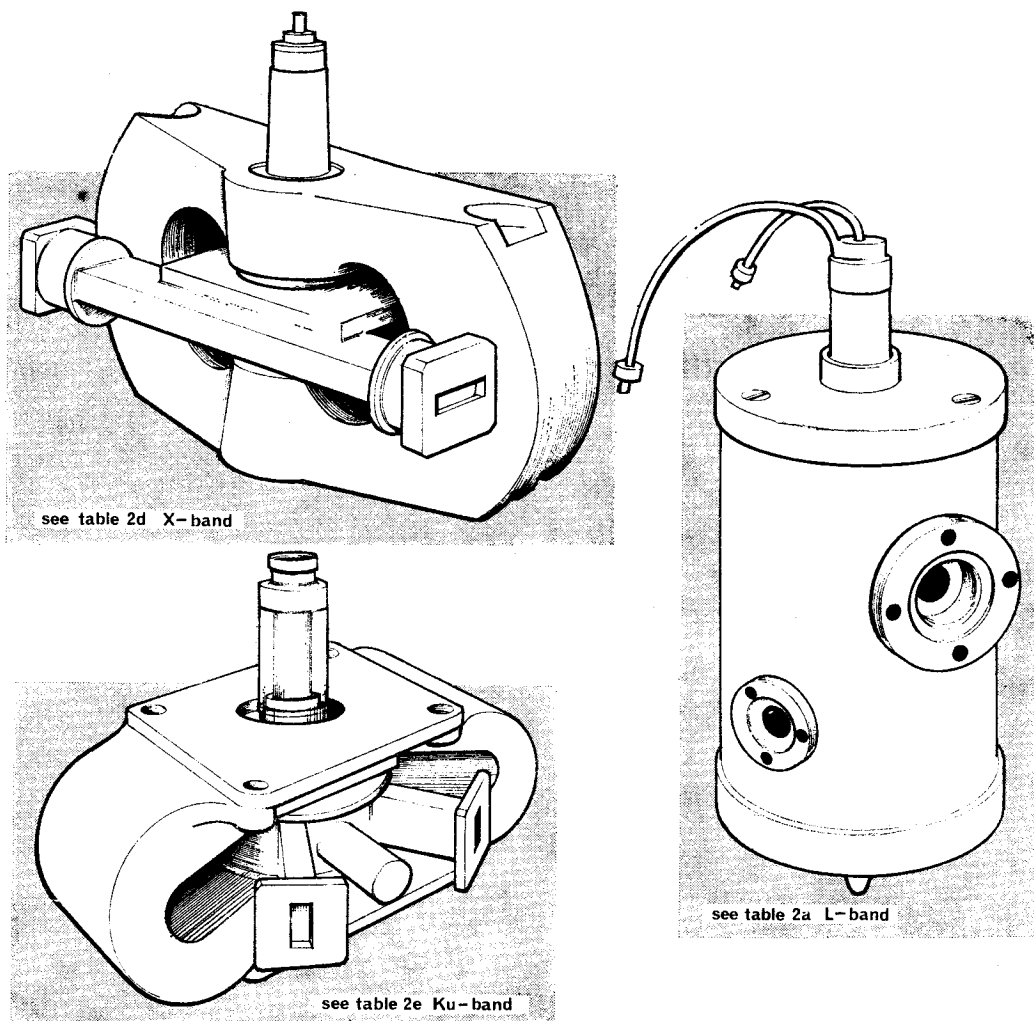


FIG. 9. TYPICAL CROSSED-FIELD AMPLIFIERS

Microwave Valve Oscillators

There are two main classifications of microwave valve oscillators: the O-type oscillator, of which the reflex klystron and the backward wave oscillator are typical; and the crossed-field or M-type oscillator, typified by the magnetron and the M-type carcinotron. In the next few paragraphs we shall examine the developments that have taken place in these devices.

Magnetron Oscillator

The magnetron (see p269) has generally been regarded as a high peak power oscillator of high efficiency, but of rather poor frequency and phase stability, necessitating the use of a.f.c. and coho circuits in associated receivers. Until recently, it was also limited in its average power capability and was confined to low duty factor operation, i.e. short pulse durations, low p.r.f.s and low mean powers. Magnetron development has been aimed at improving the frequency stability and increasing the duty factor.

Conventional pulsed magnetrons are available for operation at frequencies from 400MHz to 100GHz. They may be fixed-tuned or they may be capable of being tuned mechanically over a limited frequency range. Peak power outputs range from 100W to 5MW. Magnetrons with pulse durations up to $7\mu\text{s}$ are available, and duty factors can be as high as 0.05 (although 0.001 is more typical). The size and weight of pulsed magnetrons vary considerably, depending upon frequency of operation and available power output. Fig. 10 gives some idea of the range of variation.

Magnetrons are also available for *c.w. operation* at frequencies from 250MHz to about 5GHz. CW power outputs available range from 250mW to about 500W.

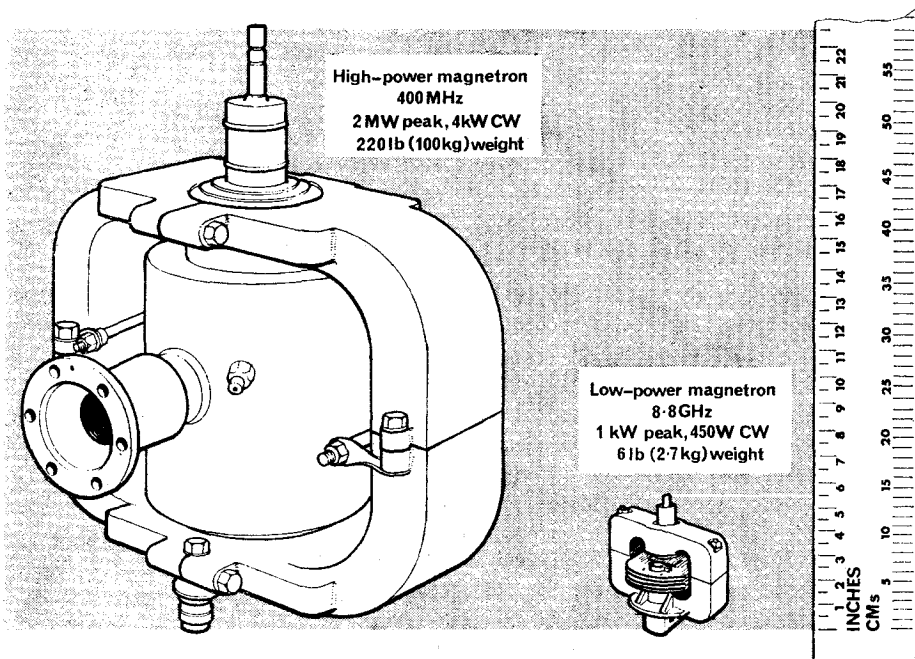
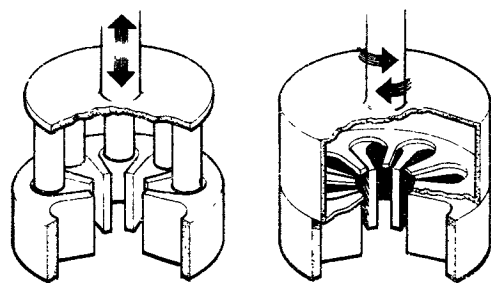


FIG. 10. VARIATIONS IN SIZES OF CONVENTIONAL PULSED MAGNETRONS

In an attempt to improve the frequency stability of magnetrons, a new design of 'coaxial' magnetron has been developed. Coaxial magnetrons have a built-in stabilizing coaxial cavity

coupled to the resonant anode structure of the valve. This high-Q cavity reduces any tendency to frequency drift which is caused by variations in temperature or load conditions.



(a) Tuning fingers

(b) Tuning disc

FIG. 11. MECHANICAL TUNING OF A MAGNETRON

As stated earlier, there is often the need to be able to alter the operating frequency of a magnetron. This may be done over a limited bandwidth, typically 5–10% of the centre frequency, by mechanical means. Fig. 11 illustrates the two usual methods. In Fig. 11a the frequency is adjusted by inserting *tuning fingers* into the anode cavities, thus altering the inductance of the cavities. In Fig. 11b a *tuning disc* with slots cut in it can be rotated round the anode block. As the disc rotates it varies both the inductance and capacitance of the anode cavities, producing a frequency sweep across the tuning range. In a typical X-band (10GHz) magnetron, a tuning bandwidth of 500MHz can be obtained by using a tuning disc.

The ability to tune a magnetron mechanically has led to the development of the *dither-tuned magnetron*. This is a magnetron in which the frequency is caused to swing backward and forward (dither) about its centre frequency of operation. If the dither-tuning is rapid enough to produce successive transmitted pulses at *different frequencies*, returns caused by sea and ground clutter are greatly reduced. This provides radar resolution of targets that would otherwise be masked by clutter. Thus, the *frequency agility* given by the dither-tuned magnetron improves target resolution, detection probability, and radar range. To reduce the clutter significantly two requirements have to be met:

- a. The magnetron must be capable of being tuned rapidly enough to produce a frequency separation between successive pulses equal to the *reciprocal* of the pulse duration at the desired radar p.r.f. Thus, if the radar has a pulse duration of $1\mu\text{s}$ and a p.r.f. of 1,000 p.p.s., the frequency difference between pulses must be at least 1MHz and the magnetron must be capable of producing this shift in 1 millisecond.
- b. From 20 to 30 separate pulses should be included within each aerial scan to cancel the clutter. Thus, in the example above, about 25 pulses, each separated by 1MHz, are required within each scan. The magnetron must have a tuning bandwidth of at least 25MHz, and be capable of tuning over this bandwidth in 25 milliseconds.

The rapid tuning requirement of the dither-tuned magnetron may be obtained by connecting the shaft of the tuning disc illustrated in Fig. 11b to a high-speed motor mounted outside the valve envelope. A typical X-band dither-tuned magnetron operates with a dither about the centre frequency of $\pm 20\text{MHz}$; it has a peak power output of 200kW, operates with a duty factor of 0.001, and weighs 12 lb. (9.5kg.).

One other important magnetron development is in '*voltage-tunable c.w. magnetrons*'. The voltage-tunable magnetron is a microwave oscillator that delivers c.w. output power efficiently over a wide frequency range; the output can be varied in frequency extremely rapidly as the *voltage to the anode* is varied. This ability to vary the frequency of the magnetron in a manner similar to that of the reflex klystron means that the c.w. magnetron can now be used for many applications that were previously denied to it.

Narrow-band voltage-tunable magnetrons (often referred to as v.t.m.s) are available from 250MHz to 8GHz; there are also *wideband* units available up to 5GHz. The c.w. output power delivered by the magnetron ranges from 30mW to 1kW.

A typical narrow-band tuning range is 20% of the operating frequency; the wideband units can have a tuning range as high as 3 : 1. Efficiency falls as the electronic tuning range is increased: at 20% tuning ranges the efficiency of the valve is of the order of 75%, and at tuning ranges of 3 : 1 the efficiency reduces to about 15%. A schematic outline of a voltage-tunable magnetron and its power supply is shown in Fig. 12. The *anode* supply controls the *frequency*, whilst the *control* electrode is an *amplitude* control.

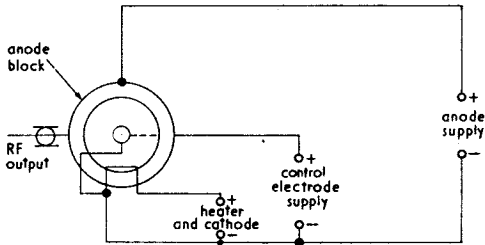


FIG. 12. POWER SUPPLY ARRANGEMENTS FOR VOLTAGE-TUNABLE MAGNETRON

A typical voltage-tunable magnetron is illustrated in Fig. 13. This magnetron is electronically-tuned over the range 3–3.5GHz as the anode voltage is varied from 1kV to 2kV. The c.w. power output is 10kW. Voltage-tunable magnetrons are finding increasing application as local oscillators in receivers, in radar altimeters, and in small f.m.c.w. transmitters. Because of its much greater efficiency and its ability to deliver a higher power output, the voltage-tunable magnetron is tending to replace the reflex klystron for many applications.

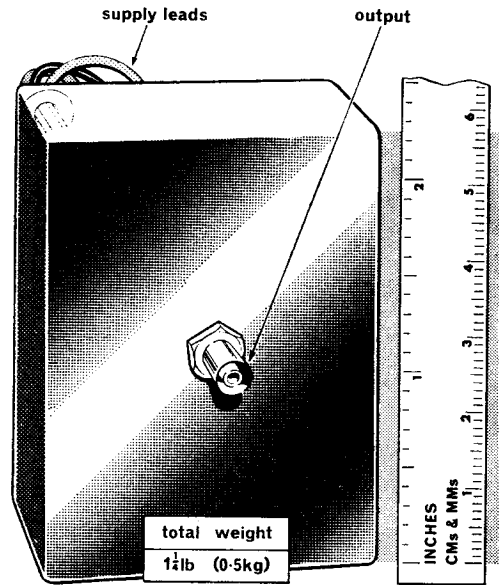
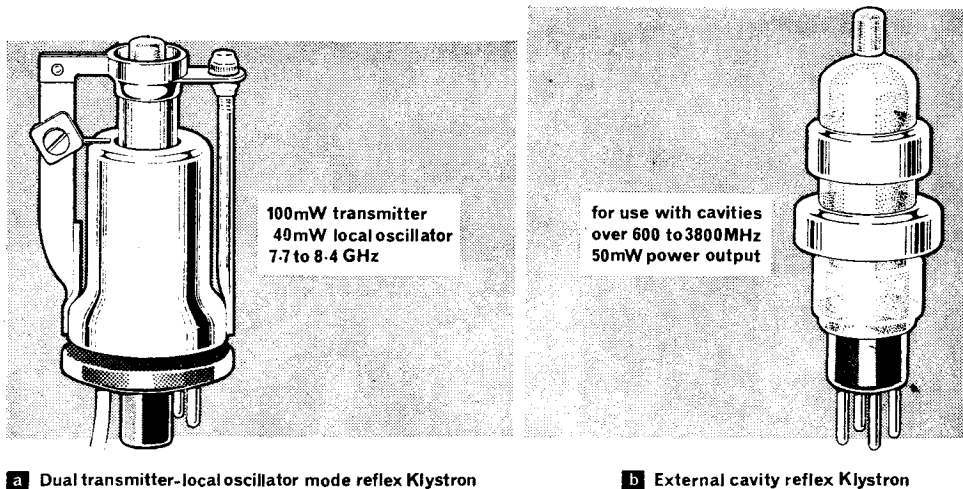


FIG. 13. TYPICAL VOLTAGE-TUNABLE MAGNETRON

Reflex Klystrons

We considered the reflex klystron in p258. No new technology has been developed for the reflex klystron, but there has been a steady improvement over the years in reliability.



a Dual transmitter-local oscillator mode reflex klystron

b External cavity reflex klystron

FIG. 14. TYPICAL REFLEX KLYSTRONS

Modern klystrons operate at higher frequencies and powers, and provide a longer operating life (now about 15,000 hours). Internal-cavity reflex klystrons are available at frequencies in the range 3.5 to 170GHz; external-cavity klystrons are available from 500MHz to 11GHz.

Available c.w. power outputs range from 15mW to about 5W, although special two-cavity oscillators (see p 256) can provide up to 300W output. Most reflex klystrons have an electronic-tuning capability, by variation of reflector voltage, typical tuning ranges being about 5% of the centre frequency. Their main use is as the local oscillator in microwave receivers, but they have also found application as low power transmitters in microwave repeaters, and in radar altimeters. Like the planar triodes mentioned earlier in this chapter, reflex klystrons have been developed to operate in a dual transmitter/local oscillator mode (Fig. 14).

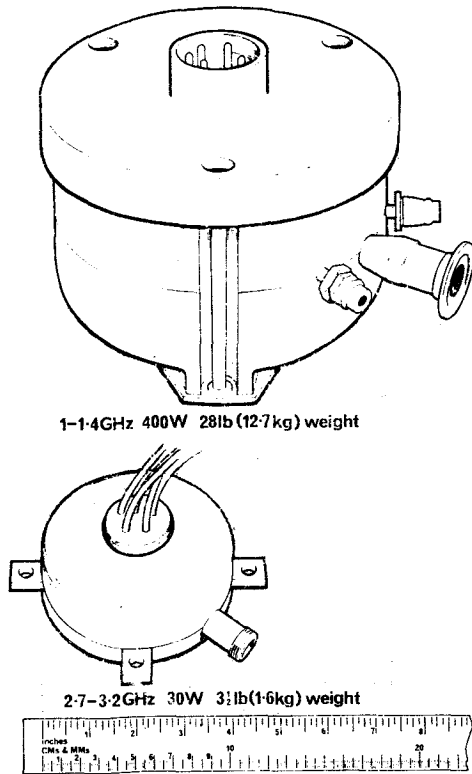


FIG. 15. TYPICAL BACKWARD-WAVE OSCILLATORS

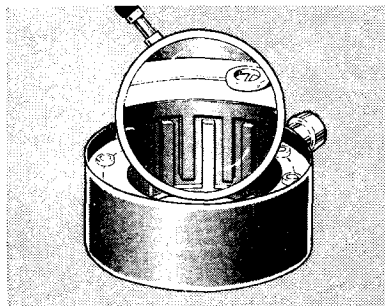
versions may be used in microwave e.c.m. transmitters.

In many modern backward-wave oscillators, a significant reduction in size and weight has

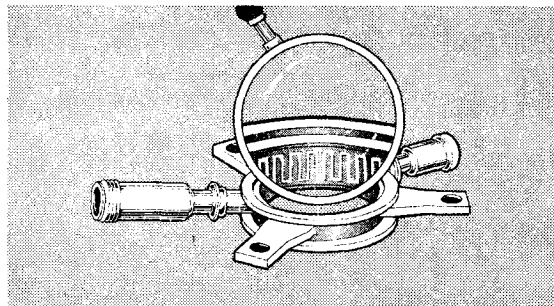
Backward-wave Oscillators

Before the advent of the voltage-tunable magnetron, the backward-wave oscillator (p265) was the only available low-power microwave source in which the frequency could be varied electronically over a *wide range*. The backward-wave oscillator has an electronic tuning range of about 40% of the centre frequency compared with about 5% for a reflex klystron.

Backward-wave oscillators are available at frequencies in the range 300MHz to 300GHz, and power outputs range from 5mW to 500W (Fig. 15). Such oscillators are reliable in operation, their noise output is low, and their frequency stability is good. The relatively high efficiency of about 40%, coupled with the wideband electronic tuning capability makes the backward-wave oscillator suitable for application as a local oscillator in wideband microwave receivers. The higher power



Conventional slow-wave structure



Ceramic-mounted thin-film slow-wave structure

FIG. 16. USE OF THIN FILMS IN VALVE CONSTRUCTION

been achieved by using thin-film techniques. Fig. 16 illustrates how conventional bulky and heavy slow-wave structures may be replaced by ceramic thin-film versions.

FERRITE DEVICES

Introduction

A brief introduction to ferrites is given on p302. In general terms we can say that a ferrite device is a component that uses the interaction between a magnetized ferrite material and an incoming signal to modify the incoming signal in certain defined ways. Ferrite devices include circulators, isolators, and gyrators. Another ferrite device that has recently been developed and that promises to have considerable application is the yttrium-iron-garnet (y.i.g.) electronically-tunable filter. The use of ferrite devices has now become so widespread that, before we consider the advances that have been made, it may be helpful to recall a few important facts about ferrite theory.

Ferrites

We know that an atom consists basically of a nucleus with several electrons orbiting at various energy levels around it (p355). Each electron can be considered as an electric charge, not only rotating about the nucleus, but also *spinning on its own axis*. This creates a magnetic field which has a direction dependent upon the direction of spin. Electrons will tend to pair themselves with electrons that have spins in the *opposite* direction, as shown in Fig. 17a. With paired electrons the magnetic field is confined to the vicinity of the electrons; with a single unpaired electron (Fig. 17b) the magnetic field is much more widespread.

For the majority of materials, paired electrons occur quite naturally and such materials exhibit little magnetic effect. However, ferromagnetic materials have atoms containing a number of *unpaired* electrons in the *outermost* orbits and this accounts for the high magnetic effect of materials such as iron and nickel. Unfortunately, these metals also have *low resistivity* so that in any interaction with an r.f. field, high current would flow giving *high loss*.

Ferrite materials are similar to ferromagnetic materials in many respects, and they also exhibit the high magnetic effect for the reason given above. There is one major difference between the two classes of materials, however, and this is the important point: ferrite materials have *high resistivity* and introduce *little r.f. loss*.

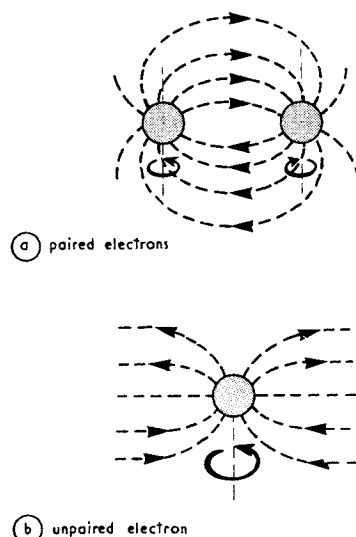


FIG. 17. ELECTRON SPIN AND RESULTANT MAGNETIC FIELDS

Action of Ferrite in Microwave Fields

Fig. 18a overleaf recalls what happens to the single unpaired electron in a ferrite material when a steady magnetic field is applied in the direction shown.

The electrons spin about their own axes but they also *precess* about the axis of the stationary magnetic field. The *direction* of precession reverses if the magnetic field is reversed and the

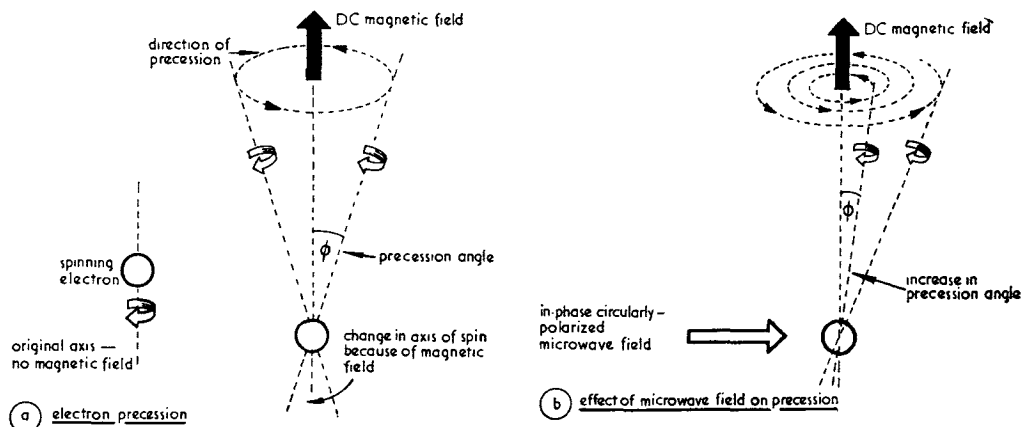


FIG. 18. ELECTRON PRECESSION AND EFFECT OF MICROWAVE RF FIELD

frequency of precession depends upon the *magnitude* of the applied magnetic field. The frequency is usually in the microwave region.

If microwave energy of this *same frequency* and of the *correct polarization* is applied to the ferrite, the electronic gyromagnetic action interacts with the microwave field, and microwave energy is absorbed by the ferrite. This is indicated in Fig. 18b which shows that when the microwave energy is 'in phase' with the precessional motion of the electron the precession angle increases, indicating that energy has been extracted from the microwave field.

Fig. 18b indicates that interaction between a microwave field and ferrite occurs only when the microwave signal has a *circularly polarized component*. In a rectangular waveguide propagating the H_{01} mode (p283) there are positions in the waveguide where the microwave magnetic field exhibits circular polarization.

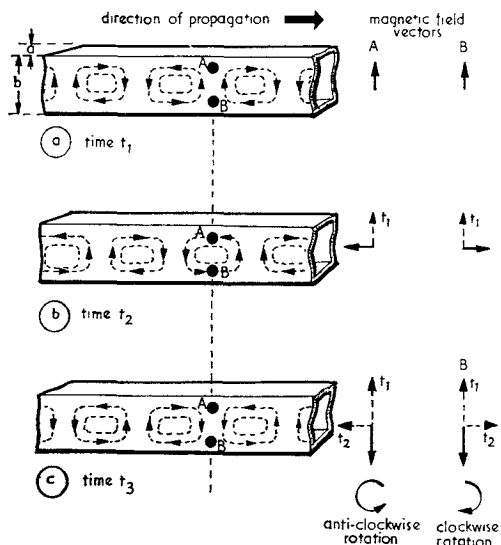


FIG. 19. POSITIONS OF CIRCULAR POLARIZATION IN A WAVEGUIDE

Consider the waveguide magnetic field pattern shown in Fig. 19. As the pattern moves down the guide, the direction of the microwave field intercepting the points A and B varies with time as shown. In effect, the field at point A rotates in an *anti-clockwise* direction, whilst that at B rotates in a *clockwise* direction. Thus, by placing a ferrite material at positions in the waveguide corresponding to points A and B, the necessary interaction between the propagating microwave field and the ferrite can be achieved.

Assuming that the ferrite is placed at the correct position in the waveguide, the ferrite will absorb energy only when the magnitude of the applied magnetic field is such that the frequency of precession coincides with the frequency of the microwave field. We thus have a form of *ferrite resonance*.

Fig. 20 shows that for a given frequency of microwave signal there is a small range of applied magnetic field strength over which microwave energy is absorbed by the ferrite. The distance between the 3db points on the ferrite resonance curve is referred to as the *linewidth*. For most applications a narrow linewidth is required.

If the position of the ferrite in the waveguide is correct for absorption of energy from an *anti-clockwise* circularly polarized field at that point, it is clear that there will be little absorption from a *clockwise* circularly polarized field, because the directions of rotation for the microwave field and the precessing electron are now *opposite*. Similarly, it may be seen that by reversing the direction of the applied magnetic field we can change the state of the ferrite from an absorbing to a non-absorbing condition, because the rotation of the precessing electron is being reversed. Thus, by changing the conditions of either the external magnetic field or the propagating microwave field we can control the attenuation of the microwave signal.

The Y-axis of the graph in Fig. 20 could also be labelled 'permeability', as the permeability of the ferrite also varies as shown in the graph. The velocity of propagation of a signal through a material is proportional to the permeability of the material, so that by varying the permeability of the ferrite the *time* taken for the signal to propagate through the material is varied. Hence the ferrite introduces *phase shift*. The amount of phase shift introduced by the ferrite depends on its magnetic biasing point on the graph of Fig. 20 and also on the length of ferrite. Furthermore, the permeability is different for clockwise and anti-clockwise circularly polarized signals, for reasons similar to that given above, so that the phase shift introduced by the ferrite is *different* for the two directions of propagation.

We shall see that most ferrite devices depend for their action on either ferrite resonance or ferrite phase shift.

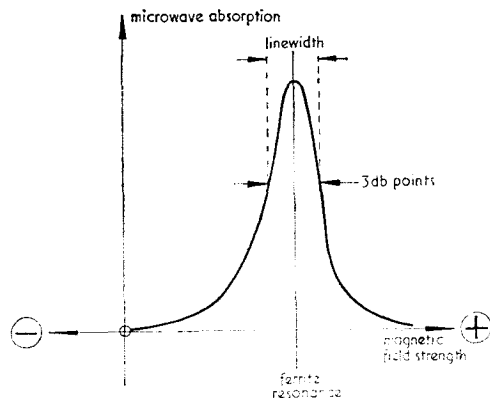


FIG. 20. FERRITE RESONANCE

Circulators

The purpose of a circulator is stated in p302: it is a non-reciprocal device having three or more ports and it is used to transmit r.f. energy from one of its ports to an adjacent port while decoupling the signal from all other ports.

There are three main types of circulators: the Y-junction circulator, the differential phase shift circulator, and the Faraday rotation circulator. The most common is the Y-junction circulator (Fig. 21). The differential phase shift circulator is used mainly for high power applications and the Faraday rotation circulator for the higher microwave frequencies above about 50GHz.

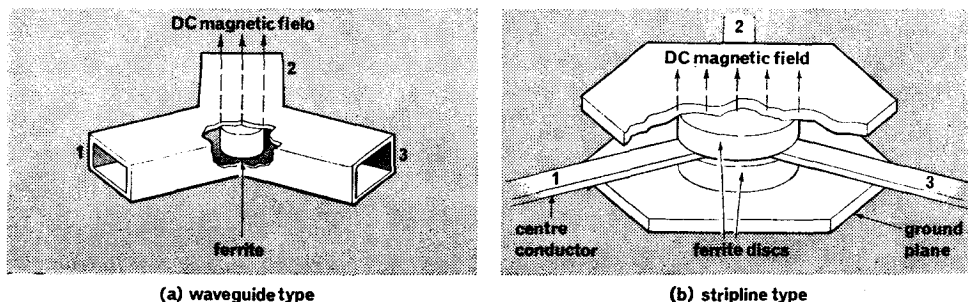


FIG. 21. Y-JUNCTION CIRCULATOR CONSTRUCTION

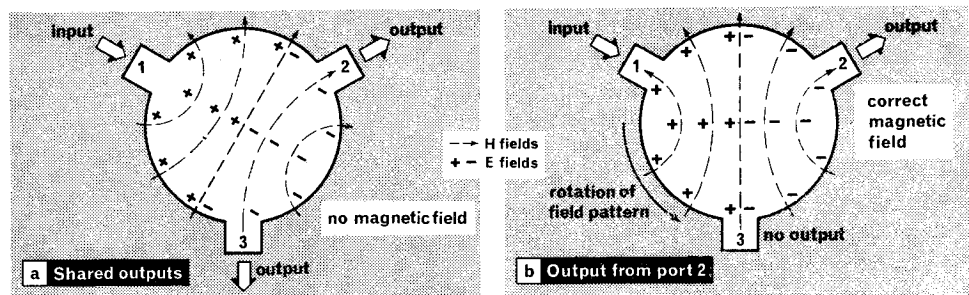


FIG. 22. BASIC Y-JUNCTION CIRCULATOR OPERATION

Y-junction circulator. This type can be constructed in either rectangular waveguide or stripline. An H-plane waveguide Y-junction circulator is shown in Fig. 21a; E-plane circulators are also available. Fig. 21b shows the stripline version, used at the lower microwave frequencies. Stripline circulators may be made with coaxial connectors.

In the Y-junction circulator, a ferrite element is placed in the centre of three junctions that are spaced 120° apart. Circulator action is obtained by magnetically biasing the ferrite along its axis with a steady magnetic field of the *correct magnitude*. The direction of circulation may be reversed by reversing the direction of the applied magnetic field.

Fig. 22 explains the basic action. In Fig. 22a the applied magnetic field is zero. When a microwave signal is applied to port 1 it is shared equally between ports 2 and 3, both outputs being 180° out of phase with the input. When the external magnetic field is applied the standing wave pattern in the ferrite *rotates anti-clockwise* as shown in Fig. 22b. For the correct magnetic bias the microwave field pattern is such that there is no output from port 3, practically all the input being available as the output from port 2. Similarly, it may be shown that a signal applied to port 2 provides an output from port 3, port 1 being isolated. Also a signal applied to port 3 gives an output from port 1, and port 2 is isolated.

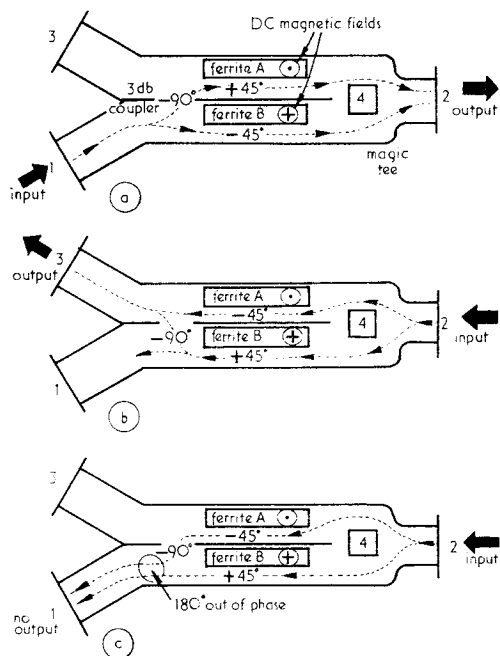


FIG. 23. DIFFERENTIAL PHASE SHIFT CIRCULATOR, BASIC ACTION

Differential phase shift circulator. A schematic outline of a four-port differential phase shift circulator is illustrated in Fig. 23. In Fig. 23a, a signal fed into port 1 is split into two parts by the 3db coupler (p313) in such a way that the energy in the top half of the waveguide *lags* that in the bottom half by 90° . The positions of the ferrite slabs and the directions of the two external magnetic fields, are such that ferrite A introduces a $+45^\circ$ phase shift while ferrite B introduces a -45° phase shift. Thus the energy from port 1 is phase-shifted in the top half of the waveguide by $-90^\circ + 45^\circ = -45^\circ$, and in the bottom half by -45° . The two portions therefore arrive *in phase* at port 2 to give the required output.

A signal entering port 2 (Fig. 23b) is split equally by the magic tee in the top and bottom halves of the waveguide. The portion in the bottom half is *advanced* by 45° by ferrite B (signal propagating in *opposite* direction) and is then divided into two equal parts by the 3db coupler; the coupled portion towards port 3 is *delayed* by 90° . The total phase shift of the energy from port 2 to port 3 via the *bottom* half of the waveguide is thus $+45^\circ - 90^\circ = -45^\circ$. The energy in the *top* half of the waveguide from port 2 to port 3 is *delayed* by 45° by ferrite A. Hence the two portions of the signal arrive *in phase* at port 3 and provide the required output.

It may also be seen that energy from port 2 to port 1 via the *bottom* path is phase-shifted by $+45^\circ$, while in the *top* path it is phase-shifted by -45° plus -90° (by 3db coupler) to give a total phase shift of -135° . The two portions are therefore 180° *out of phase* and cancel (Fig. 23c).

Similarly, it can be shown that a signal applied to port 3 will appear only at port 4, and one applied at port 4 will appear only at port 1. Thus we have a four-port circulator.

Faraday rotation circulator. A Faraday rotation four-port circulator is illustrated in Fig. 24.

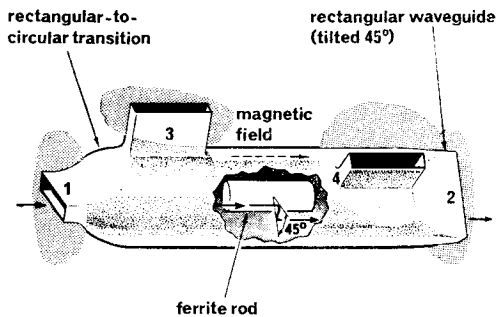


FIG. 24. FARADAY ROTATION CIRCULATOR, BASIC ARRANGEMENT

This device depends for its operation on the fact that a linearly-polarized wave can be considered to be made up of two circularly-polarized waves of equal magnitude rotating in *opposite* directions (see earlier). If a ferrite rod is inserted along the centre line of a circular waveguide and an external magnetic field is applied along the axis of the rod, the two circularly polarized components of the microwave field are shifted in phase *differentially*. This is because one wave *aids* electron precession in the ferrite whilst the other *opposes* it. The resultant polarization is thus *rotated* about the axis of the ferrite rod, either clockwise or anti-

clockwise depending on the direction of the applied external magnetic field. The *amount* of rotation depends on the length and diameter of the ferrite rod; increasing either will increase the amount of rotation.

In Fig. 24 the H_{01} rectangular mode signal applied to port 1 is converted to H_{11} circular mode in the circular guide where it interacts with the ferrite rod. The signal is rotated 45° clockwise by the ferrite rod and passes as the output to port 2. This signal affects neither port 3 nor port 4 because the electric fields do not cut these ports. A signal applied to port 2 is acted upon by the Faraday rotator to produce an output in port 3 and no other. Similar remarks apply for ports 3 and 4 so that circulator action is achieved.

Comparison of circulators. Circulators are available for operation at all frequencies within the range 30MHz to 220GHz, the Faraday rotation type being used for the higher frequencies. They are capable of handling powers up to 150kW c.w. and 30MW peak, the differential phase shift type being used for the higher power levels. The insertion loss (e.g. from port 1 to port 2) is typically of the order of 0.5db; and the isolation (e.g. in the direction port 2 to port 1) is usually about 25db.

As we have seen, ferrite circulators can be constructed in various forms (see Fig. 25 overleaf). One of their main applications in radar is as a TR switching device to prevent the large transmitted power damaging the sensitive receiver. In general, in a three-port circulator, if the transmitter is connected to port 1, the aerial to port 2, and the receiver to port 3, the conditions necessary for satisfactory TR switching are met.

Until recently in c.w. radars it was often necessary to provide *separate* aerials for transmitter and receiver to give adequate isolation. The improvement in ferrite circulators at the higher power levels has now to a large extent overcome this limitation. The result is that a single

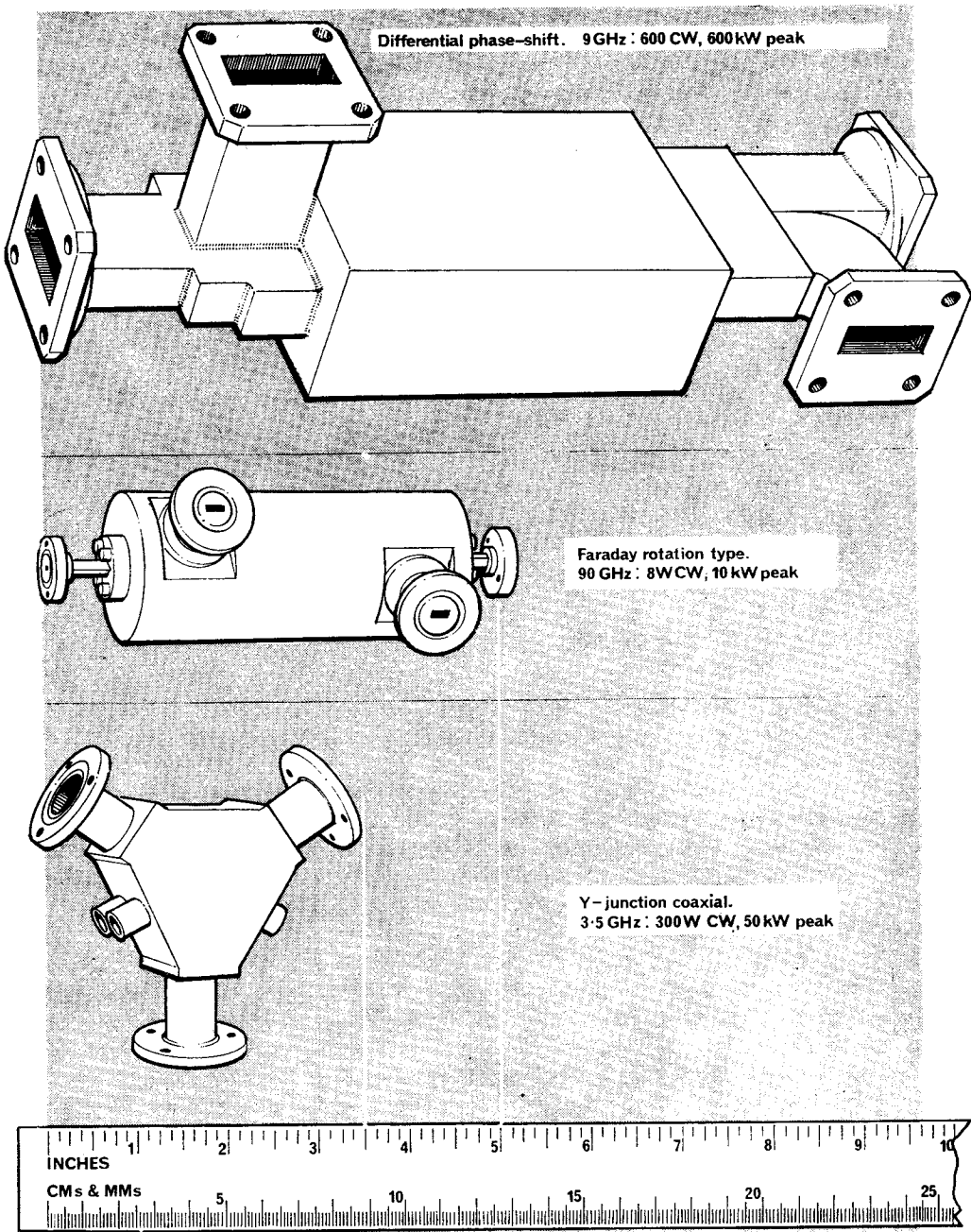


FIG. 25. EXAMPLES OF MODERN CIRCULATORS

aerial connected to a suitable circulator may now be used for those high power c.w. radars that previously needed separate receiver and transmitter aerials.

Isolators

We saw on p 302 that an isolator transmits a signal in one direction with little loss of energy whilst a signal in the other direction is heavily attenuated. There are three main types of ferrite isolators: the resonance isolator, the terminated-circulator type of isolator, and the Faraday rotation isolator.

Resonance isolator. We saw earlier that there are positions in a rectangular waveguide where the microwave field exhibits circular polarization. If a thin ferrite slab is located at such a position in the waveguide (Fig. 26) and it is biased with a steady magnetic field of the correct magnitude to give ferrite resonance the ferrite gives *directional absorption*. With the biasing magnetic field and the propagation of the microwave signal in the directions indicated, the precessing electrons in the ferrite rotate in the same direction as, and in phase with, the microwave magnetic field in the waveguide. The microwave signal is therefore *heavily attenuated* as the ferrite absorbs microwave energy. For a microwave signal propagating in the *other direction* very little absorption occurs and the signal passes with little loss. This is because the microwave magnetic field is now rotating in the *opposite* direction to that of the precessing electrons.

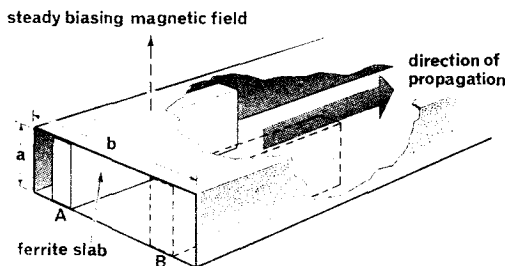


FIG. 26. OUTLINE OF RESONANCE ISOLATOR

Note that if the ferrite slab is moved from position A to position B, the *direction of isolation* is reversed. Similar remarks apply if the direction of the biasing magnetic field is reversed, the ferrite remaining in position A.

Terminated-circulator isolator. This type of isolator is simply a three-port Y-junction circulator with port 3 terminated with a matched load to make a *two-port* device. Energy is transmitted with little loss from port 1 to port 2, but is heavily attenuated in the reverse direction. The terminated circulator can usually handle higher powers than the resonance isolator and is often smaller.

Faraday rotation isolator. A Faraday rotation isolator is illustrated in Fig. 27. The action is

similar to that of the Faraday rotation circulator discussed on p585. A ferrite rod is inserted along the centre line of a circular waveguide and suitably biased by an external magnetic field acting along the axis of the rod. As we have seen, this effectively rotates the microwave polarization about the axis of the rod; in this case a rotation of 45° is used. The forward wave is unaffected by the resistive element because the electric field is perpendicular to it. After being rotated by 45° the forward wave is at the correct angle to pass to the output port. The

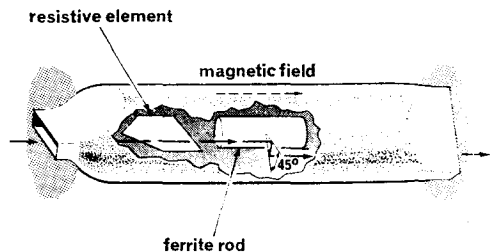


FIG. 27. FARADAY ROTATION ISOLATOR, BASIC ACTION

reverse signal, however, is rotated by the ferrite so that the electric field is *parallel* to the resistive element and is *attenuated*. This gives the required isolation. The Faraday rotation isolator is used only at the higher microwave frequencies.

Comparison of isolators. Isolators are available for operation at all frequencies within the range 30MHz to 200GHz. They are capable of handling powers up to 150kW c.w. and 10MW peak, the terminated circulator being used for the higher power levels. The insertion loss is typically of the order of 0.5db and the isolation about 25db.

There are many occasions in radar where it is necessary to allow energy to pass from one point A to another point B with little loss, whilst prohibiting the passage of energy in the direction B to A. Ferrite isolators satisfy these conditions. Fig. 28 illustrates typical examples of available ferrite isolators.

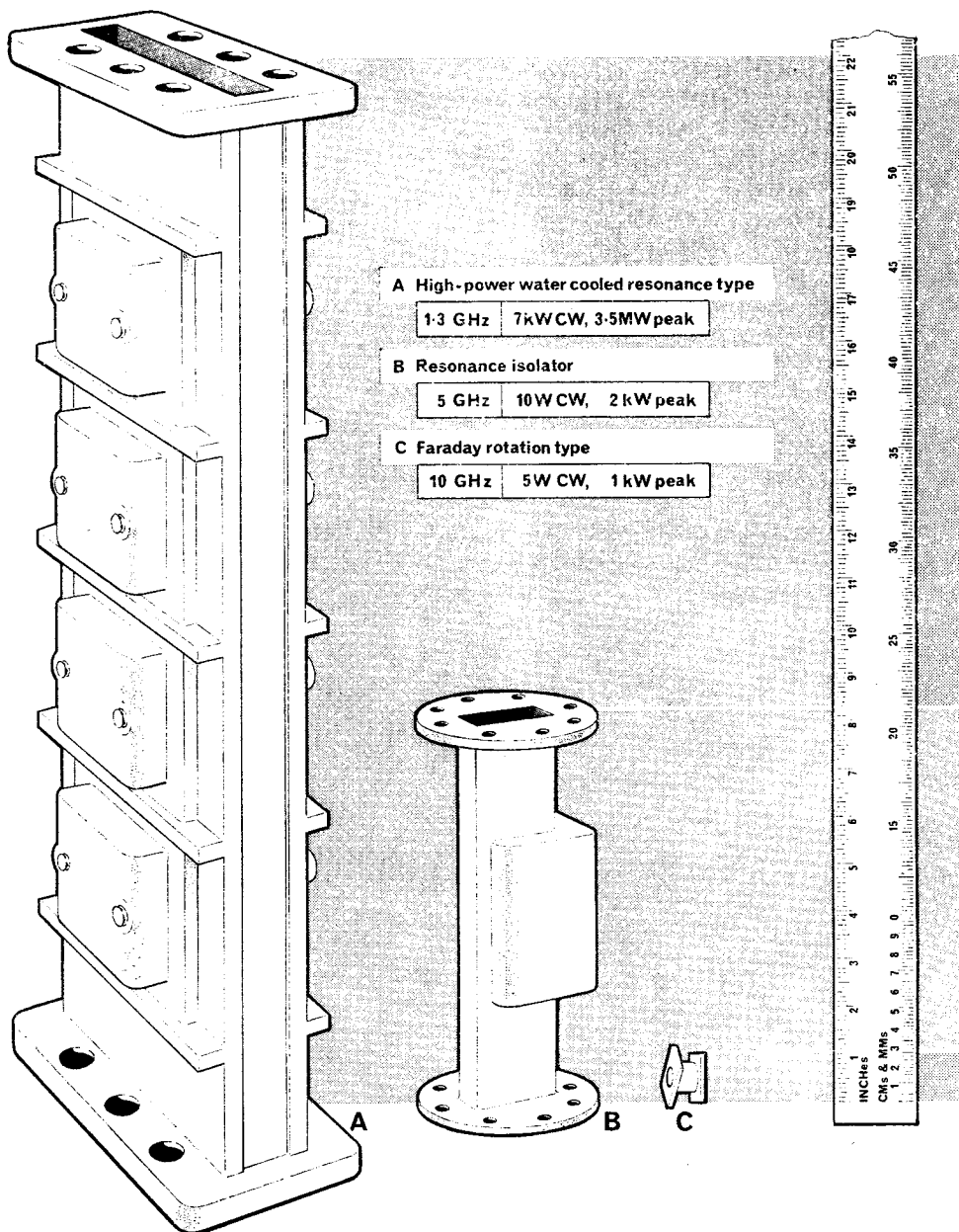


FIG. 28. EXAMPLES OF CURRENT FERRITE ISOLATORS

Ferrite Gyrotors and Phase-shifters

Ferrite gyrotors are considered very briefly on p302, where one type is illustrated. They are used to introduce a differential phase shift such that a microwave signal transmitted in one direction undergoes a phase shift relative to a signal transmitted in the other direction. The gyrotor normally introduces 45° , 90° , or 180° differential phase shifts.

We saw earlier that if a magnetically-biased slab of ferrite is placed at a point of circular polarization in a waveguide, the permeability of the ferrite is different for the two directions of propagation. The differential phase shift introduced is then proportional to the *difference* in permeabilities. The schematic outline of a typical gyrotor is shown in Fig. 29a.

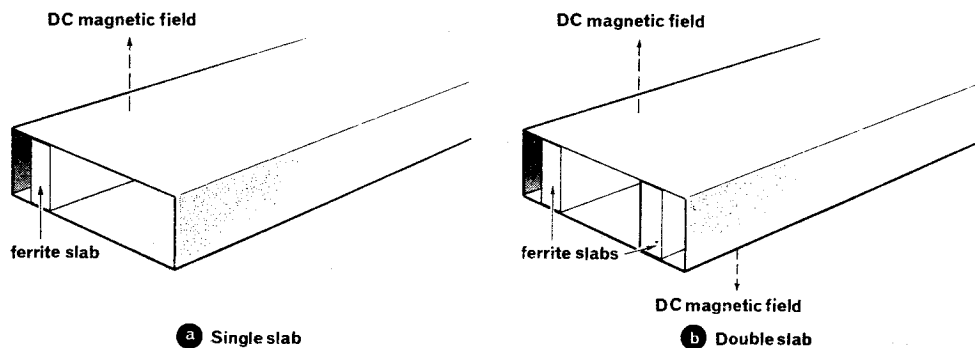


FIG. 29. OUTLINE OF FERRITE GYROTORS

The similarity to the resonance isolator (p587) may be noticed. The important difference between the two devices is in the *value of magnetic bias*: in a gyrotor, the magnetic field is *less* than that required to produce ferrite resonance.

For a given ferrite material the *amount* of phase shift depends on the value of the external magnetic field and also upon the length of the ferrite slab. To increase the differential phase shift without increasing the length of the device, *another* slab of ferrite which is *oppositely* magnetized may be inserted at a point of *opposite* circular polarization in the waveguide (Fig. 29b). With both the external magnetic field and the direction of circular polarization changed the phase shift introduced by the second slab is in the *same direction* as that in the first, and the two phase shifts *add*.

Digital phase-shifters. Gyrotors are used to introduce a *fixed value* of phase shift in a given direction of transmission. There is also a need for a device in which both the value and direction of phase shift can be *varied* quickly and easily. The digital phase shifter provides this. A basic digital phase shifter is illustrated in Fig. 30a.

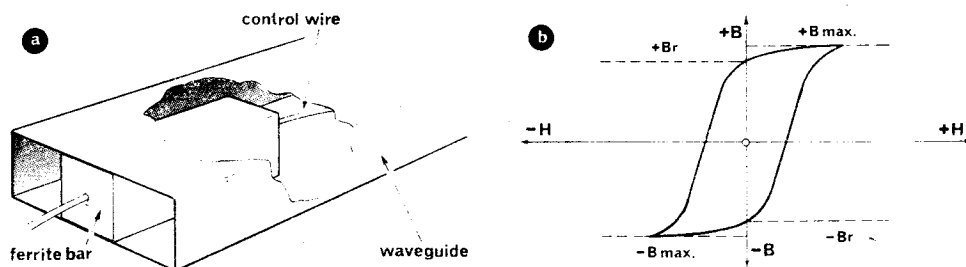


FIG. 30. BASIC LATCHING DIGITAL PHASE-SHIFTER AND OPERATION

It consists essentially of a ferrite bar placed in the centre of a waveguide, with a control wire running through the centre of the ferrite. The ferrite material used for this has pronounced *hysteresis*, as shown by the curve of Fig. 30b. Thus if a positive pulse of current is passed through

the control wire sufficient to magnetize the ferrite to $+B_{\max}$. the ferrite remains magnetized to $+B_r$ when the pulse is removed. Such a device is termed a 'latching' device because it requires no holding current.

On the application of a current pulse the ferrite becomes oppositely magnetized on *either side* of the control wire; that is, if the magnetic field on the left of the control wire is in an 'upward' direction, that to the right of the wire is 'downward'.

Hence the arrangement is similar to that of the conventional two-element gyrator illustrated in Fig. 29b. The microwave signal interacts with each side of the ferrite bar to give a large 'positive' phase shift. If we now apply a *negative* pulse of current to the control wire, the magnetism latches at a value of $-B_r$ when the pulse is removed. For the same direction of microwave propagation we now get a *negative* phase shift.

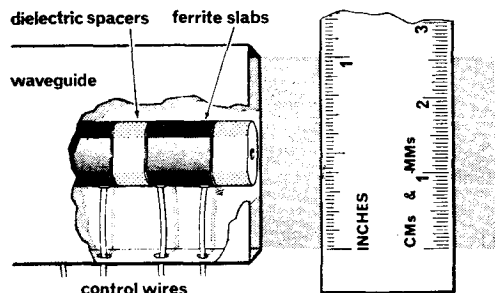


FIG. 31. MULTI-ELEMENT DIGITAL PHASE-SHIFTER

pulsed and can be computer-controlled if required. Its main application to date is as a fast electronic phase-shifter for phased aerial arrays (electronically-steered arrays). It is available at all frequencies within the range 800MHz to 150GHz. Phase shifts of 360° can be obtained at switching speeds of less than $1\mu s$, and at drive powers of less than 1W. Peak powers of up to 100kW can be handled.

YIG Filters

We noted earlier that the resonant frequency of a piece of ferrite placed in a waveguide may be changed by varying an applied external magnetic field, and for a given frequency of microwave signal the magnetic field may be adjusted to give ferrite resonance. The *size* of the ferrite in the devices so far discussed is very much larger than the wavelength of the microwave signal, so that at resonance the microwave power absorbed by the ferrite is converted to heat within the crystal structure and is not available for transfer.

However, if we have a sphere of ferrite material of such a size that it is *comparable* with a half-wavelength of the microwave signal, electromagnetic resonances occur in the small spherical crystal. Such a sphere, suitably inserted in a waveguide or coaxial cavity, acts as a *resonant circuit*. It has the advantage over other types of resonant circuit that its resonant frequency may be easily varied over a wide frequency range by *varying the applied magnetic field*.

Therefore in the digital phase-shifter we have a device which, by the application of a single pulse of current of the correct polarity, provides the required change in phase. For a given material and microwave frequency, the phase change depends on the *length* of the ferrite. By connecting elements of any required length in cascade, as in Fig. 31, and switching each element as required, a latching digital phase shifter, having as many discrete values of phase shift as desired is obtained.

The digital phase-shifter requires no holding current and no magnet; the control current is

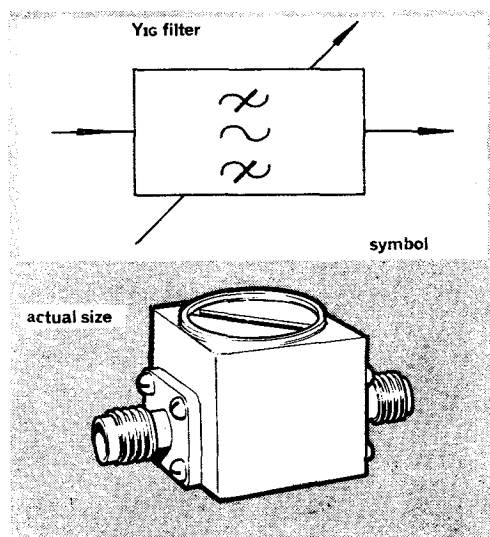


FIG. 32. TYPICAL YIG FILTER

To be suitable as a filter, a very narrow linewidth, i.e. high Q , is required. A suitable material is available in *yttrium-iron-garnet* (y.i.g.). A properly prepared y.i.g. crystal of the correct (very small) wavelength dimension provides an effective filter of high Q .

A typical y.i.g. filter is illustrated in Fig. 32.

Electronically-tunable y.i.g. filters are available at frequencies within the range 90MHz to 50GHz. The tuning ratio is of the order of 2 : 1 (e.g. tuning from 2 to 4GHz). The 3db bandwidth is typically 30MHz, but this may be reduced by using two or more y.i.g. spheres in cascade. The insertion loss is about 3db, and the isolation 'off tune' is about 30db. The resonant frequency varies with the applied magnetic field, and the tuning sensitivity is quoted as so many MHz per mA magnetizing current; a typical sensitivity is 4MHz/mA. The y.i.g. filter can handle only very low powers, up to about 10mW peak. However, its applications do not demand high power-handling capability. Its main use is in tuning the r.f. and local oscillator stages of microwave receivers.

DELAY LINES

Introduction

There are many occasions in radar where it is necessary to *delay* a signal for a short period of time, usually of the order of microseconds. We saw an example of this in the delay line canceller for m.t.i. radar (see p502).

The ultrasonic, or acoustic, delay lines considered on p505 operate satisfactorily at frequencies up to about 100MHz. If we require to delay a *microwave* signal the usual procedure is to convert the signal to a *lower* frequency at which the delay line can operate efficiently. The lower frequency signal is then delayed in the normal manner and then reconverted to its original form at a later time. This is the system used in the m.t.i. delay line canceller mentioned above.

For some applications, however, conversion to a lower frequency in order to achieve delay is not practicable, e.g. where the bandwidth of a modulated microwave signal is very large and exceeds any convenient lower frequency. In such cases it becomes necessary to examine ways in which a microwave signal may be delayed *directly*. This implies that delay devices capable of operation at microwave frequencies are available. In the following paragraphs we shall examine developments in such devices.

Superconducting Delay Lines

Until recently, the only way of conveniently delaying a microwave signal by direct means was to pass the signal through a length of coaxial cable or waveguide. With conventional coaxial cable, to achieve a delay of $3\mu\text{s}$ at a frequency of 3GHz, a cable length of about 2,000 feet (610 metres) would be required and this would weigh about 750 lb. (340kg.). Probably even more important is the fact that such a length of cable would have an attenuation of over 1,000db at this frequency. From this it may be seen that the two aims in developing suitable delay devices are *reduction in physical size* and *reduction in losses*. Whilst the use of waveguide in the above example would reduce losses, the resulting delay line would be huge.

One method, which has been developed to achieve small size and low loss is the use of *superconducting* materials, such as niobium, in *miniature* coaxial cables. The miniature cable can be coiled up compactly so that it occupies a very small volume, and by operating the cable at a suitably low temperature, at which the resistance of the cable is negligible, the attenuation may be reduced to a few decibels. A typical superconducting miniature cable designed to introduce a $2\mu\text{s}$ delay has an insertion loss of only 2db at 1GHz, rising to 8db at 10GHz. The only disadvantage of this system of delay is the need to operate at *very low temperatures*, approaching Absolute Zero.

Electron Beam Delay Lines

The basic outline of an electron beam microwave delay device is illustrated in Fig. 33. An electron gun forms the beam which then passes through an input coupler where it is modulated by the input signal. After being delayed in the delay section of the valve, the electron beam is demodulated by the output coupler, and the delayed microwave signal passes to the output.

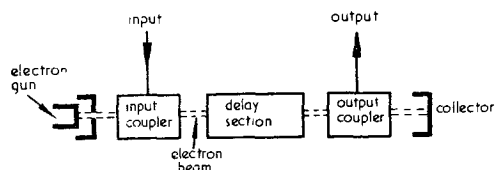


FIG. 33. OUTLINE OF ELECTRON BEAM DELAY DEVICE

delay of the order of $10\mu\text{s}$ the tube would need to be 60 yards (55m.) long. To reduce the length of the delay section, the beam velocity must be drastically reduced. The obvious way of doing this is to reduce the beam voltage. However, at very low beam voltages, the electron beam becomes unstable and subject to deflection by stray electric and magnetic fields.

To achieve low beam velocity, and hence small tube length, at reasonable voltages, the principle of crossed d.c. electric and magnetic fields is applied. An electron moving in that shown in Fig. 34. The exact shape of the path depends upon the initial velocity of the electron, but the *forward velocity* in the direction a-b-c is *low* and is dependent only upon the values of the d.c. electric and magnetic fields. Thus the delay introduced may be *varied* by changing the values of the d.c. fields.

An electron beam delay device based on the principle described here, and measuring only 9 in. (23cm.) long, has been designed to introduce an $8\mu\text{s}$ delay at a frequency of 2GHz. The delay may be varied as stated above and the signal attenuation is only about 6db. The main disadvantage of this device is that it is not solid state.

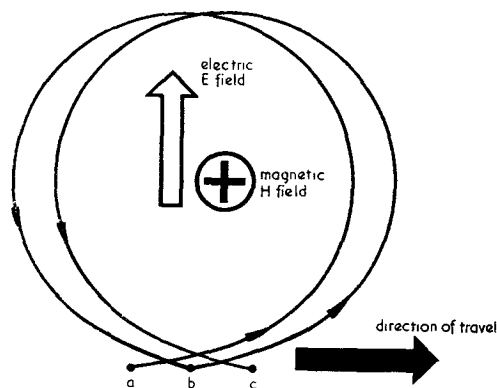


FIG. 34. ELECTRON BEAM IN CROSSED DC FIELDS

Acoustic Delay Lines

Acoustic delay lines, typified by the quartz delay line discussed in p506, have been used at frequencies up to about 100MHz for many years. Basically, the electromagnetic input signal is converted by a transducer to an acoustic wave which travels through the delay line medium at a velocity of only a few thousand metres per second, as opposed to a free space velocity of 3×10^8 metres per second. At the far end of the line the acoustic energy is converted back into electromagnetic energy by another transducer. Because of the relatively low velocity of acoustic waves, considerable delays can be obtained in short lengths of suitable material, e.g. a $1\mu\text{s}$ delay can be obtained in a quartz delay line less than $\frac{1}{2}$ in. (12.7mm.) long.

To make this device suitable for operation at *microwave* frequencies, lower loss delay line materials and improvements in transducer design are needed. By using materials such as sapphire, rutile, and single-crystal y.i.g. as the delay line medium, very low losses at microwave frequencies have been reported.

Semiconductor and thin film techniques have been applied in an attempt to solve the *transducer* problem and good electromagnetic-acoustic conversion efficiencies have been obtained using these techniques, but much development work remains to be done in microwave acoustic delay lines. Available units designed for use at 10GHz provide large delays with losses of less than 20db.

It has also been found possible to obtain *amplification* as well as delay from acoustic devices. In this way some of the losses associated with the delay line may be offset. For amplification, a *piezoelectric* semiconductor material is used as the delay medium. Fig. 35 illustrates the basic idea. A voltage is applied across the semiconductor to produce a *drift of electrons* along the bar. At the same time, a small microwave signal applied to the left end of the bar is converted into an acoustic signal by the input transducer. Because of the piezoelectric effect, the acoustic wave is accompanied by electric fields travelling along the bar at the same velocity as the acoustic waves. The electric fields associated with the acoustic waves interact with the drifting electrons in a manner similar to that of a travelling-wave tube. The result is that energy is extracted from the electrons and the acoustic energy increases. After conversion by the output transducer we obtain a *delayed* microwave signal of *greater amplitude* than the input. Gains of up to 100db have been reported.

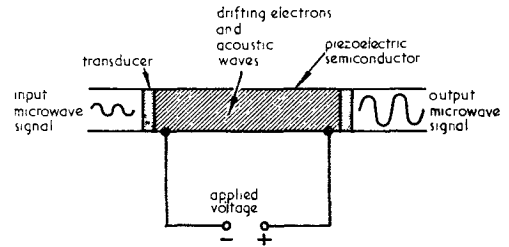


FIG. 35. ACOUSTIC AMPLIFICATION

Ferrite Delay Lines

We have seen that a y.i.g. crystal may be used as the delay medium in an *acoustic* delay line. If, at the same time, the ferrite is placed in a suitable magnetic field, *electron precession* takes place in the ferrite. The acoustic waves then interact with the precessing electron field in such a way that the available delay may be *varied* by varying the applied magnetic field. The low weight, small size, and variable delay capability of the ferrite delay line make it suitable for several applications in radar.

The potential of this device has not yet been realized fully, mainly because of the difficulty of producing suitable transducers. Practical units available at the present time include a device with $3\mu\text{s}$ delay and a loss of only 15db at 1GHz. Efforts are being made to reduce the transducer losses still further.

Summary

In this chapter we have considered developments that have taken, and are taking, place in important radar components. In a book of this type only the most important points about the more common devices can be sketched in. We cannot deal with every new component. However, the trend for the future in radar devices and components is indicated in this chapter and in the previous one. Where such components are being introduced into RAF equipments, further details will be found in the appropriate Air Publication.

DEVELOPMENTS IN AIRBORNE RADAR SYSTEMS

Introduction

Airborne radar equipments are many and varied. Some equipments are common to many different types of aircraft; other equipments are more specialized and designed for use in fighter, bomber, transport or reconnaissance aircraft. Although aircraft do not carry all the equipments mentioned below, the 'common' items may be taken to include:

- a. **Airborne Doppler radars.** See p477.
- b. **Radar altimeters.** See p465.
- c. **IFF equipment.** This is a secondary radar transponder carried by all aircraft to provide identification to an interrogating ground radar station (see later).
- d. **Range-measuring system.** A range-measuring system is a secondary radar equipment providing a pilot with the distance and the bearing of his aircraft to a selected ground beacon. An example of the system is the Rebecca-Eureka equipment mentioned briefly on p50. A more modern equipment (known as TACAN) provides greater ranges and more accurate bearings. Air-to-air TACAN is also in use, enabling aircraft to rendezvous with other aircraft.
- e. **Hyperbolic navigation systems.** Three types of hyperbolic navigation systems are in use: Gee, Loran, and Decca. Such systems enable an aircraft to fix its position by means of two or more intersecting position lines generated by ground beacons. The airborne equipment consists simply of a receiver and a suitable display. Each position line is established by comparing the received transmissions of two ground beacons. The position lines obtained in this way take the form of families of *hyperbolae*. A second family of hyperbolae can be obtained by comparing the transmissions from *another pair* of ground beacons. Thus, it is possible to make measurements to obtain *two* position lines whose intersection provides a position fix for the aircraft. The comparison between the transmissions from the pair of ground beacons may be on the basis of a *time difference* between pulses, as in Gee and Loran; or it may be on the basis of a *phase difference*, as in the Decca system. A summary of the theory of hyperbolic systems is given in AP1234C.

In addition to the common items of radar equipment, specific tasks can be accomplished efficiently only with the aid of special radar equipments. Examples include:

- a. **Airborne interception equipment (AI).** AI is fitted in fighter aircraft to assist in the final stages of interception of enemy aircraft (see p48).
- b. **Navigation and bombing system (NBS).** NBS is a self-contained airborne navigation and blind bombing aid used in bomber aircraft. It contains an X-band search radar which is used to produce a p.p.i. map of the ground over which the aircraft is flying. The radar output is also applied to a computer where, in conjunction with other inputs, the track, groundspeed, and position of the aircraft are computed. This information may be applied to the automatic pilot to steer the aircraft to a specific target. When over the target, the computer automatically initiates the release of the bomb.
- c. **Anti-surface vessel radar (ASV).** This equipment is carried in marine reconnaissance aircraft for the detection of surface vessels and submarines (see later).

It is not proposed to deal further in this book with the equipments listed above. Although some of the equipments are complex, necessitating special training, all of them use combinations of the principles described earlier in this book. Specific details of equipments are given in the appropriate Air Publications.

In this chapter we wish to examine newer radar systems which, in many cases, use principles different from those associated with conventional pulsed or c.w. radars. Many of these newer equipments are already in use in the RAF and it is, therefore, important for us to know something about them.

Airborne Reconnaissance Systems

With the advent of supersonic aircraft and ballistic missiles it is vital for this country to be able to obtain early warning of an enemy's intention. One of the most effective ways of doing this is by means of airborne reconnaissance. There are two main categories of reconnaissance, demanding different techniques:

- a. Surveillance of the ground in certain areas, requiring accurate maps and pictures of the territory flown over, for study and interpretation on the aircraft's return to base.
- b. Marine reconnaissance for the detection of submarines and surface vessels, the information obtained often calling for immediate action.

For ground-mapping and surveillance, modern reconnaissance aircraft are fitted with a sideways-looking radar system, a line scan system, and photo-reconnaissance cameras. Marine reconnaissance aircraft may also be fitted with these equipments but, in addition, a powerful scanning radar of the anti-surface vessel (ASV) type is usually carried. Basic information about all these electronic reconnaissance systems is given in the following paragraphs.

Sideways-looking Radar

Introduction. Many airborne radars used for ground-mapping or surveillance are rotating aerial types, producing a conventional p.p.i. display. With such radars the scanner is usually mounted in a radome under the aircraft fuselage, and for practical reasons, it is necessary to limit the *size* of the aerial. Because of this limitation, the aerial radiation pattern tends to be *wide* in both azimuth and elevation planes so that the resolution of this type of radar is poor. Furthermore, since the radar beam is being swept in a circle with the aircraft at the centre, an enemy is given advance information of the aircraft's arrival, and electronic countermeasures can be applied.

With *sideways-looking radar* (s.l.r.) both these disadvantages are overcome. The s.l.r. looks sideways and downwards from the aircraft, with the radar beam *fixed* at right angles to the fore-and-aft axis. The aerial does not rotate and is mounted parallel to the axis of the aircraft. Because of this the aerial may be made long, so giving a *narrow* azimuth beamwidth and good resolution in this plane. The forward motion of the aircraft itself is used to scan the ground beneath. Since the radar beam is directed to the *side* of the aircraft it avoids giving advance information of the aircraft's arrival to an enemy's countermeasures.

With such a radar the ground on either side, or both sides, of the aircraft is scanned by the aircraft's forward motion, and a picture of the ground may be built up *line by line* on photographic film.

Basic sideways-looking radar. The basic elements of a sideways-looking radar are a short-pulse microwave transmitter, a wideband receiver, a long aerial, and a photographic recorder. The only moving parts of the system are the aircraft itself and the recording film.

The aerial radiates a beam which is fan-shaped in the vertical plane and *very narrow* in the horizontal plane. The equipment is usually designed for *two aeriels*, one looking to port and the other to starboard, the radar being switched from one to the other on successive pulses or for groups of pulses. In this way we can produce two maps of parallel strips of ground spaced equally either side of the aircraft (Fig 1).

The beam intercepts the ground some distance to the side of the aircraft flight path. Each pulse of energy illuminates a strip of ground extending in a direction at right angles to the aircraft

heading. Such an area can be anything from one to 50 miles long (A to B in Fig 1), depending upon the height of the aircraft and the angle of depression of the aerial beam; the strip may be only 50 ft (15m) wide where the beam intercepts the ground (B to C), depending upon the azimuth beamwidth of the radiation pattern. The angle of depression of the aerial beam is usually about 45° so that the width of the area immediately beneath the aircraft which is not covered by the radar is about *twice* the height of the aircraft.

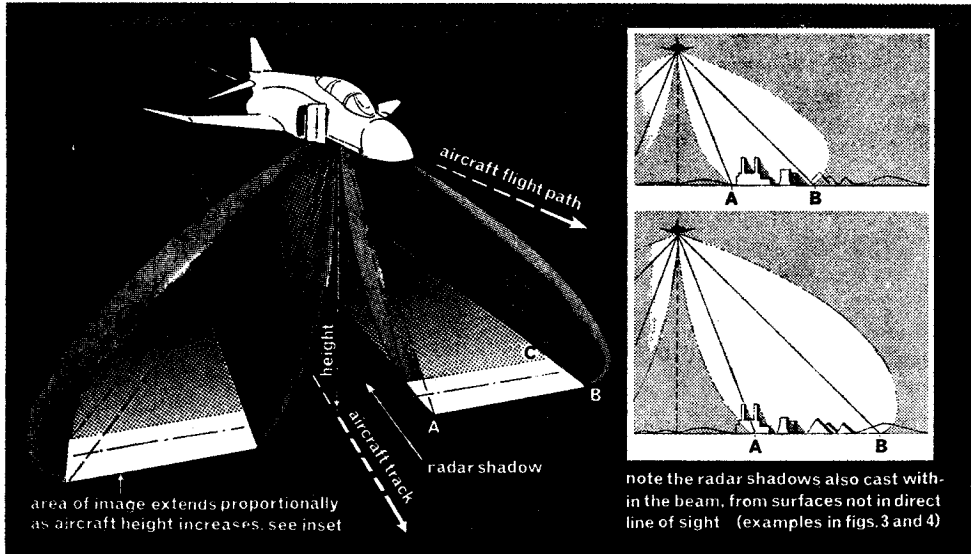


FIG 1. RADIATION PATTERN FOR SIDEWAYS-LOOKING RADAR

Every part of the ground which is intercepted by the radar beam is illuminated by many radar pulses and reflects *many echoes* before it passes out of the beam. Thus the rate at which the information is being collected by the radar is much higher than the rate at which it is changing so that the 'integration' of the radar is high. This produces a high quality picture of good resolution in contrast with that produced by a conventional scanning radar in which the information changes significantly from scan to scan, giving poor 'integration'.

Photographic recorder (Fig 2). The output from the radar receiver is used to intensity-modulate a linear timebase on a high-resolution c.r.t. The information in the ground returns is

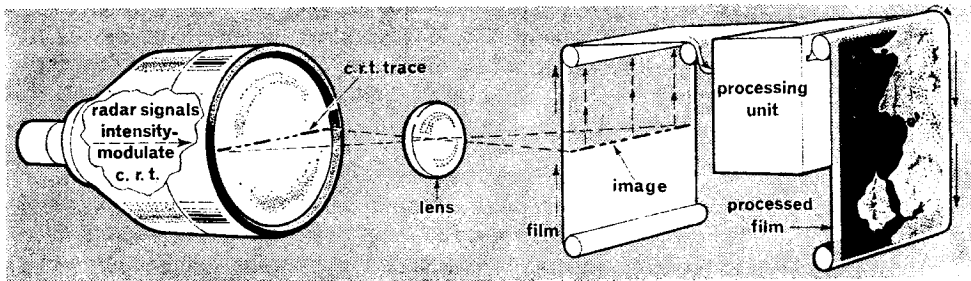


FIG 2. AIRBORNE PHOTOGRAPHIC RECORDER SYSTEM FOR SLR

thus conveyed by the *brightness* of the c.r.t. trace along its length. The length of the trace is proportional to the distance AB in Fig 1 and so represents the ground *range* intercepted by the radar beam. An image of the c.r.t. trace is projected through a lens on to film which is arranged to move continuously past the image at a rate proportional to the groundspeed of the aircraft (obtained from the airborne Doppler radar).

After exposure to the light projected from the c.r.t. the film passes under a processing head and the fully-processed film emerges within a matter of seconds for viewing in the aircraft. The film forms a complete record of what the radar has seen of the territory over which the aircraft has flown, and may be used as a navigational aid.

When a film is developed and processed quickly the picture quality suffers. If immediate decisions based on the recording have to be made, quality is of secondary importance. But if an accurate map is required for survey or intelligence, the resolution must be improved. It is then necessary to process the exposed film much more slowly under controlled conditions on the ground after the aircraft has returned to base. The film can then be viewed by skilled interpreters. In such cases the processing part of the airborne photo recorder is not used.

The quantity of film required for any operation depends upon the height at which the aircraft is flying. For example, a 50-miles survey from a height of 1000 ft (300m) would need 45 ft (13.5m) of film; from a height of 500 ft (150m), *twice* as much film would be needed to cover the same area.

Picture quality. Sideways-looking radar produces a picture which has good definition at both high and low altitudes, and the fidelity of the reproduction is high. Unlike aerial photography, sideways-looking radar can produce a plan range map of good quality in *all weathers*, through cloud and in the dark. Fig 3 illustrates a mosaic map of part of the southern coastline of England produced by sideways-looking radar from high altitude.

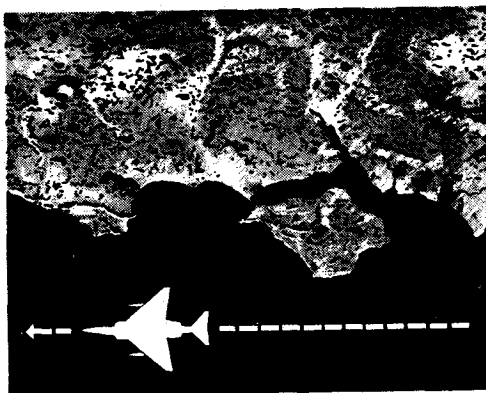


FIG 3. TYPICAL HIGH-LEVEL SLR MAP

At heights of 500 ft (150m) or less, vertical surfaces tend to reflect a stronger signal than horizontal ones. The strong signal returned by a vertical surface is further emphasized by the 'shadow' area immediately behind it, from which no radar returns are received. The result is that the picture tends to become three-dimensional and it is possible to make out features which would not usually show up on a conventional radar p.p.i. display. The effect is exaggerated when the aircraft is flying over hilly country because of the increased amount of shadow. Similarly, the effect is noticeable in built-up areas because of the large numbers of vertical surfaces. When operated over the sea the responses from surface vessels, because of their vertical sides, are in sharp contrast to the background sea return. Fig 4 illustrates the three-dimensional effect of pictures produced by sideways-looking radar at low altitudes. The differences between the radar map and the ordnance survey map may be accounted for by the fact that the radar picture was taken during high tide, when some of the sandbanks and beaches were under water.

Typical equipments. We have already seen that the aerial used in sideways-looking radar should be as long as possible to achieve a narrow azimuth beamwidth. Aerials as long as 50 ft (15m) have been used at frequencies of 10GHz and the usual design is that of a slotted waveguide array. To reduce the radiation pattern in the vertical plane the waveguide is usually placed within a horn type reflector, and to equalize the signals returned from equal-sized objects at all ranges a horn with a *cosecant-squared* radiation pattern is used (see p321).

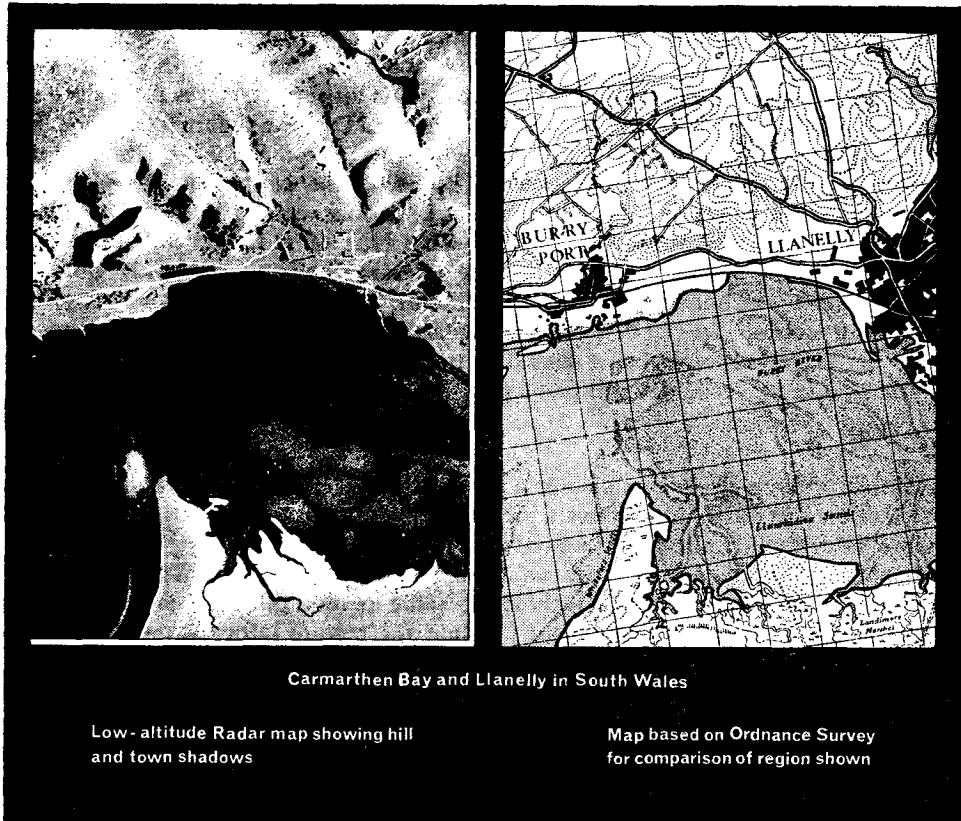


FIG 4. SHADOW GIVES THREE-DIMENSIONAL EFFECT AT LOW ALTITUDES

Short-duration pulses are required for good resolution in range, a typical value being $0.25 \mu\text{s}$. The receiver is normally a superheterodyne type with balanced crystal mixers. The receiver bandwidth must be wide enough to handle the returned short-duration echoes without distortion. The i.f. amplifiers normally have a *logarithmic* response (see p432) so that the large dynamic range of the incoming signals is compressed to a level that the film can reasonably handle.

In some equipments a height-finding facility is included and this allows radar height above the ground to be measured to an accuracy of $\pm 100 \text{ ft}$ (30m) in the range 1000 ft (300m) to 56,000 ft (17,100m).

Modern s.l.r. equipment is lightweight, very reliable, and easy to operate. Semiconductor devices are used almost exclusively. In modern aircraft the s.l.r. equipment may be fitted into a reconnaissance pod that can be attached to the fuselage (Fig 5). The equipment can be operated at any altitude up to 70,000 ft (21,300m) in an unpressurized environment, and ranges covered by the radar vary from one mile to 50 miles depending upon aircraft height.

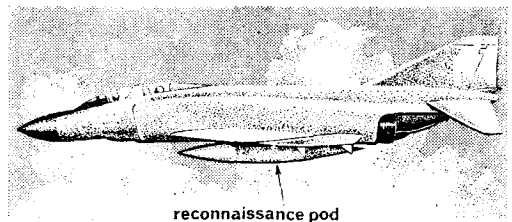


FIG 5. TYPICAL MODERN AIRCRAFT WITH RECONNAISSANCE POD

Moving-target indication. Airborne m.t.i. can be included in sideways-looking radar equipment. The basic principles are similar to those discussed for ground m.t.i. on p499. In airborne m.t.i. a radar receiver is used in which the echoes from stationary objects are suppressed and moving targets only are displayed. The problem with airborne m.t.i. is that, since the radar is in motion, *all echoes*, stationary and moving, will contain a Doppler shift in frequency. However, the Doppler shift for all stationary objects is the *same*, whereas moving targets have an *additional* Doppler shift caused by their *own movement*. By using the consistent Doppler shift from stationary objects as a *reference*, the airborne m.t.i. receiver can be designed to differentiate between moving targets and stationary objects. This type of m.t.i. (known as non-coherent m.t.i.) actually depends for its operation on the reference shift of stationary objects and is particularly suited to airborne reconnaissance radars where a continuous clutter signal is received from the ground. Special delay-line cancellers are used to remove the reference Doppler shift, the moving-target responses being passed on to the display after amplification. In sideways-looking radar, the m.t.i. responses are painted as peak signals against a mapping background.

Line Scan

Introduction. Although line scan is not a radar device, it makes use of scanning techniques, similar to those used in radar, to produce a picture of the ground over which an aircraft is flying. It is therefore worthy of mention in this book.

Taking photographs by means of air cameras from low altitudes and at high speeds presents certain problems: wide-angle coverage is difficult; some method of illumination is necessary for night photography; and if it is required to transmit the picture information from the aircraft back to base, bulky processing equipment is needed in the aircraft. Line scan, which can be designed to operate either at optical or at infra-red frequencies, overcomes these limitations.

Optical line scan. In line scan, a rotating mirror mounted under the fuselage of the aircraft scans successive strips of ground as the aircraft moves forward. The intensity of the light reflected from the ground varies for different objects, and this variation in intensity is detected by a photoelectric multiplier cell placed so that it can continuously view the mirror. The electrical output of the photocell may be used to intensity-modulate a linear trace on a high-resolution c.r.t., the length of the trace being such that it represents the length of the strip of ground being scanned. Film is moved past the c.r.t. at a rate proportional to the groundspeed of the aircraft and the exposed film then bears a record of the ground over which the aircraft has been flying. This film may be processed in the aircraft for immediate viewing; or it may be taken back to base for controlled processing and interpretation.

If immediate viewing of the record *on the ground* is needed, the electrical output of the photocell may be used to modulate an r.f. signal in a data link transmitter for *transmission* from the aircraft to the ground. The received r.f. signal, after demodulation, can then be used to produce the required picture on film.

Optical line scan may be either *passive* or *active*. In the passive role the ground being scanned is illuminated by natural sunlight. In the active role, the ground is illuminated by *artificial means* from the aircraft. In active optical line scan, two scanning mirrors are used. Both mirrors rotate in synchronism and both are directed to cover the same strip of ground together. One mirror supplies the scanning artificial light whilst the other receives the reflected light from the ground for application to the photocell.

Passive optical line scan can be used only during hours of daylight. Active optical line scan can be used at night as well. But neither system provides satisfactory results through cloud or fog.

Infra-red line scan. This method of line scan overcomes the limitations of optical line scan in cloudy or foggy conditions. The scanning infra-red detector records the *temperature differences* of objects on the ground and so provides information to produce a picture of the ground over which the aircraft is flying. Infra-red line scan requires no artificial illumination (i.e. it operates

in the passive mode) and can be used by day or night; it can 'see' objects through cloud or fog; and it can detect camouflaged objects. Although the resolution is poorer than with optical line scan, the infra-red system can pick out 'hot spots' on the ground. It is, therefore, of great tactical importance for the detection of concealed industrial targets and of any other source producing heat (e.g. the engines of military vehicles).

Range of line scan. Line scan is normally used at aircraft heights of between 200 ft (60m) and 2000 ft (600m); but in passive systems, operation up to 50,000 ft (15,000m) is possible.

At 2000 ft (600m) a line scan system will produce a picture of a strip of ground extending about one mile either side of the aircraft. At 200 ft (60m) the total width of the strip is reduced to about 400 yards (360m).

Typical line scan pictures. Typical line scan pictures are illustrated in Fig 6. Fig 6a is a picture of a group of buildings, and Fig 6b is a picture of a ship. Fig 6a shows some distortion towards the left hand side of the illustration. The reason for this is that the aircraft is operating at low altitude and the scanning mirror is looking vertically downwards in the centre of the picture but at an oblique angle at the extreme ends of the scan. By viewing the picture through a correcting lens the distortion disappears.

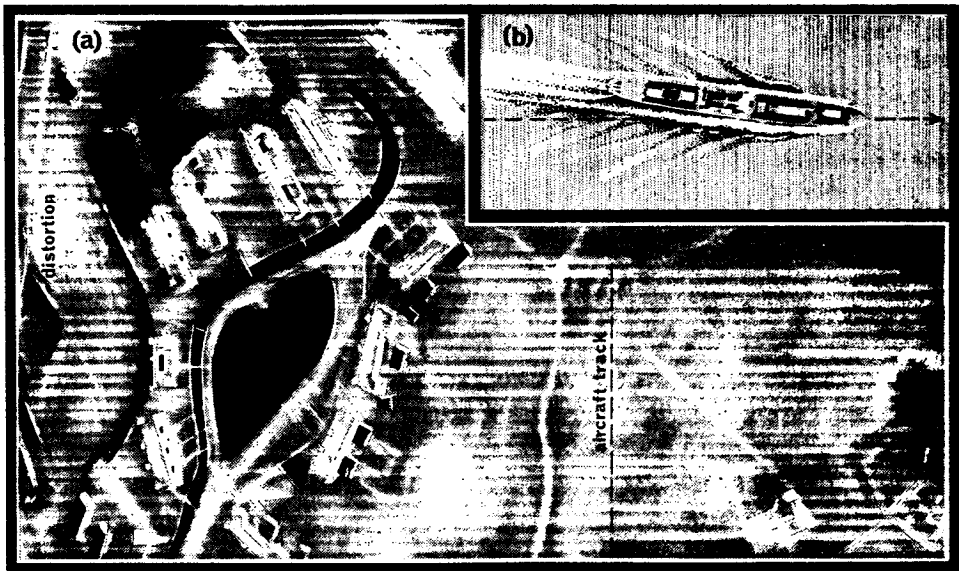


FIG 6. EXAMPLES OF PICTURES PRODUCED BY LINE SCAN

Marine Reconnaissance

In marine reconnaissance the major function is the location of submerged submarines. At present, fully-submerged submarines cannot be detected by airborne radar systems and other methods have to be used for this. On the other hand, a submarine on the surface is no more difficult to detect than a surface vessel, and even a small portion of a submarine, such as a periscope or a snort, can readily be detected by modern airborne radars.

The type of radar used for the detection of surfaced or partly-submerged submarines is usually a high-powered scanning radar of the ASV type. This type of radar has a p.p.i. presentation and has been specially designed for periscope detection. It can also pick up ships and land

masses at considerable distances so that one of its secondary functions is as a navigation aid. In addition, its scanner may be tilted to provide cloud or collision warning.

One of the limitations of a high-powered pulsed radar scanning ahead of the aircraft is that its transmissions can be intercepted by a submarine's e.c.m. equipment at almost twice the range at which the aircraft can itself detect the submarine. This allows ample time for the submarine to dive and escape detection.

This limitation may be overcome by using sideways-looking radar. The detection ranges obtainable are less than with the scanning radar but sideways-looking radar gives no advance warning to a submarine. A further advantage of sideways-looking radar is that its resolution is good so that snorts or periscopes can be detected in high seas. Another possibility is the use of line scan, but at the moment this has limited range.

Radar and other scanning systems are useful for submarine detection only when the submarine shows some portion of itself above the water. With modern nuclear-powered submarines this does not happen very often. Because of this, non-radar detection devices have been developed. Such devices include:

- a. Sonobuoys.* These are buoys that are dropped from the aircraft into the sea to act as stationary sonar sets, listening for under-water noises. The noises picked up modulate the sonobuoy radio transmitter which passes the bearing information of the noise source to the aircraft. By using two or more sonobuoys a position fix on a submarine can be made. The main limitation is in the number of sonobuoys that can conveniently be carried by the aircraft.
- b. Magnetic anomaly detector.* This device measures small local changes in the earth's magnetic field. For submarine detection it works on the principle that any metal object causes a distortion of the earth's magnetic field which can be measured. The range is limited because the aircraft has to fly very low to get any useful reading at all.
- c. Gas detection.* Some marine reconnaissance radars carry a device for sensing that a diesel-driven submarine had been on the surface in that area shortly before the aircraft arrived. Of course, the detected gas could equally be from a diesel-driven surface vessel; and a nuclear-powered submarine produces no diesel exhaust gases.
- d. ECM analysis.* ECM detection depends upon the submarine emitting some form of radio transmission. An aircraft fitted with modern e.c.m. direction-finding equipment may then be able to locate the source of transmission.
- e. Infra-red detection.* A submarine raises the temperature of the water around it and a sensitive infra-red device in the aircraft could then detect the change in temperature and record it on film, as in infra-red line scan.

Terrain-following Radar

Introduction. From previous work we know that a ground-based, long-range search radar tends to be ineffective at low angles of elevation. At such angles the clutter from ground returns may be so great that low-flying targets escape detection. Attacking aircraft can take advantage of this limitation by flying 'under the radar screen'.

However, flying at such low altitudes also brings problems to the pilot of the attacking aircraft. With modern high-speed aircraft, avoiding obstacles by means of visual observation of the ground ahead of the aircraft's track is an almost impossible task. Hence, some *automatic* means must be found for informing the pilot of obstacles ahead so that he has ample time in which to take the necessary avoiding action. This is the function of the airborne terrain-following radar. It provides aircraft with a means of following the contours of the ground ahead so allowing attacking aircraft to operate safely at low heights, under the radar screen, by day or by night and in conditions of poor visibility.

The basic equipment consists of a parabolic reflector, a transmitter-receiver, a computer, and power supplies. As we shall see later, the computer is also supplied with other inputs. The equipment operates in the microwave band (10GHz, 13GHz, or 24GHz). The transmitter provides a peak pulse output of several kilowatts at a high p.r.f. (typically 3000 pps.) and short pulse duration (typically 0.5 μ s).

Basic principles. The terrain-following radar is a forward-looking pulsed radar which detects obstacles ahead of the aircraft and supplies 'fly-up' or 'fly-down' signals for display on suitable instruments. The pilot can then take the necessary avoiding action based on the display. Alternatively, the output from the radar may be applied to the automatic pilot for control of the aircraft flight path.

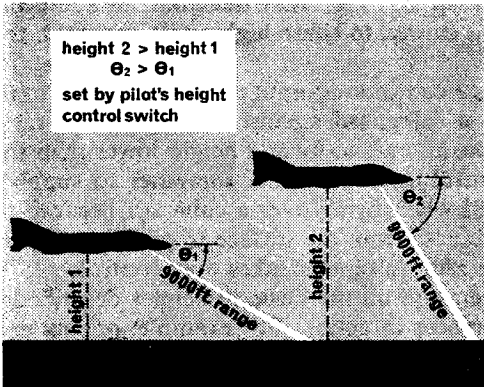


FIG. 7. VARIATION OF AERIAL DEPRESSION ANGLE WITH HEIGHT, TERRAIN-FOLLOWING RADAR

These include:

- A drift angle input from the airborne Doppler radar to align the scanner axis with the aircraft track.
- An input from an airstream direction detector; this consists of angle-of-airflow transducers fitted to either side of the nose of the aircraft to provide information on the angle of attack of the aircraft in both pitch and roll axes.
- An input from the radar altimeter as a back-up for height information.

Fig 8 illustrates the basic action of the terrain-following radar. When the aircraft is flying over flat ground at the height selected on the pilot's control panel, the radar beam intercepts the ground ahead at a slant range of 9000 ft (2700m) and the computer produces an output which is indicated on the head-up display as 'fly-level' (Fig 8a).

If now the ground ahead of the aircraft rises, the radar return gives a slant range of less than 9000 ft. The computer then produces a 'fly-up' demand which is displayed to the

The radar produces a pencil beam which points ahead of and below the aircraft; the equipment is adjusted such that the maximum available slant range is 9000 ft (2700m) so that only objects within this range provide a radar return. The pilot selects the height at which he wishes to fly and this height setting determines the *depression angle* of the parabolic reflector and hence of the radar beam (Fig 7). It is essential that the beam depression angle relative to the aircraft's flight path be maintained and this is achieved by servo control of the radar scanner. To stabilize the scanner various other inputs to the system are needed.

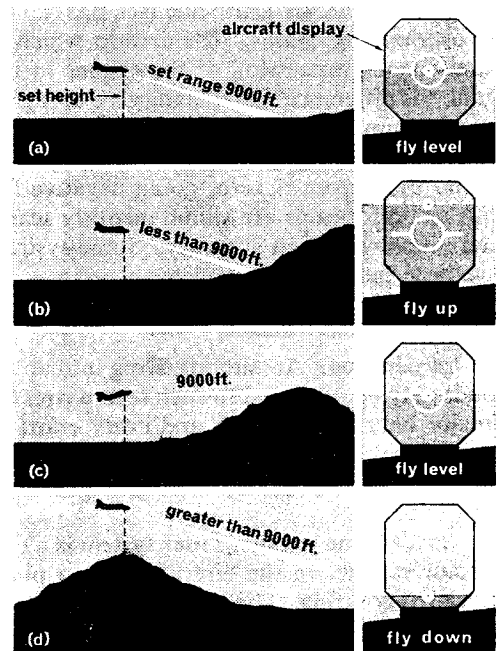


FIG 8. BASIC IDEA OF TERRAIN-FOLLOWING RADAR

pilot (Fig 8b). By pulling up, the pilot zeroes the demand, keeping the spot within the circle on the display (Fig 8c).

As the aircraft passes over the hill and the ground falls away ahead the radar beam no longer touches the ground and the computer then produces a 'fly-down' demand (Fig 8d).

By applying the appropriate control movements the pilot is able to keep the aircraft at the selected height above the ground. In doing so he follows the contours of the ground.

If the aircraft is approaching a very steep slope the *rate* at which the slant range is varying *changes rapidly* and this causes the terrain-following computer to produce a 'climb high' signal. This signal, in turn, causes the scanner to be depressed by a small angle from its controlled setting. The result is that the measured slant range is *decreased* and the radar demands an additional *increase* in pitch from the aircraft. Thus the aircraft is caused to climb high on its approach to a steep hill or a vertical surface.

As mentioned earlier the radar altimeter provides a continuous input of vertical height to the terrain-following computer. This provision acts as a safeguard against errors in slant range measurement. If the terrain-following radar allows the aircraft to fly at a height lower than that selected, the radar altimeter acts as an override control and causes the computer to supply a 'fly-up' demand. The radar altimeter is also essential when flying over a calm sea because the terrain-following radar will then receive insufficient returns from the surface to give an adequate indication of slant range. Under such conditions the radar altimeter takes over control.

Most terrain-following radars depend for their operation on the basic principles described in the previous paragraphs. There are, of course, differences in detail. For example, one modern terrain-following radar uses a scanner which scans vertically about its pre-selected depression angle. A vertical scan of $+8^\circ$ to -12° about the datum is used and ground information at all scanner angles is fed to the computer, thus producing a 'profile' of the terrain ahead of the aircraft. This profile is compared with the ideal flight path over flat ground and causes the computer to initiate the necessary fly-up or fly-down demands.

To obtain the high accuracy in range needed for terrain-following, the radar may use *monopulse* in the vertical plane (see p337). Two feeds are used to generate two slightly divergent but overlapping beams. The area in which the beams overlap is known as the *radar boresight*. Simultaneous processing of the echoes in both beams produces a sum signal and a difference signal, the difference signal being zero along the boresight. The phase and amplitude relationships between sum and difference signals are such that returns from *above* the boresight produce a *positive* output, whilst those from *below* give a *negative* output. Where the boresight intersects the ground the output is zero, going negative for a decrease in range and positive for an increase. Thus as both beams are simultaneously scanned vertically, the slant range to the ground can be determined with high accuracy. Some systems are so accurate that runs against television masts have been made with complete success.

Airborne Weather Radars

Introduction. An aircraft flying into an area of high turbulence associated with thunderstorm clouds may be subjected to such severe stresses that a fatal crash results. Such areas are obviously a major hazard to aircraft and every effort is made to avoid flying into them. It is possible to detect thunderstorm clouds by means of microwave pulsed radar, and aircraft fitted with airborne weather radar equipment are able to navigate through weather fronts and so avoid areas of turbulence.

An airborne weather radar system is a comparatively simple device. It consists of a parabolic reflector scanner in the aircraft nose, a pulsed radar transmitter-receiver, a p.p.i. display, and associated controls. The reflector is arranged to scan a sector of $\pm 90^\circ$ in azimuth with reference to the fore-and-aft axis of the aircraft. The scanner is usually stabilized by means of a gyro-controlled servo system to maintain a fixed scan plane despite roll and pitch movements of the aircraft.

Factors affecting operation. The radar echo from a thunderstorm cloud is obtained from the energy scattered back towards the radar from the particles existing within the cloud. We saw on p228 that for radar detection of an object the wavelength in use must be comparable in size with the object. Thus, for detection of particles within a cloud the radar wavelength must be very small i.e. the frequency must be *high*. On the other hand, if the frequency is too high, atmospheric absorption is high and the radar signal is greatly attenuated (p228).

Frequencies used in practice range from 3GHz to 20GHz. At low altitudes, where atmospheric absorption tends to be very high, the lower frequencies of 3GHz (S band) or 5GHz (C band) give best results. At normal jet aircraft cruising altitudes of 20,000 to 50,000 ft the most common frequency is 10GHz (X band). For very high altitudes above 50,000 ft, 20GHz (K band) has been used. The majority of current weather radar systems operate around 10GHz in X band.

The effective echoing area of a cloud is proportional to the volume illuminated by the radar pulse. This volume depends upon the width of the radar beam and also upon the transmitter *pulse duration*. By increasing the pulse duration the effective target echoing area is increased, thereby increasing the magnitude of the radar echo received back at the radar. In addition, since a long pulse duration means that we can use a narrower receiver bandwidth, the signal-to-noise ratio of the system is improved. Pulse durations in use at the moment are in the region of 2 μ s, but newer equipments are being designed with a pulse duration of 4 μ s, and it is possible that a figure of 10 μ s will be used for future equipments.

Typical equipments. The schematic outline of a typical airborne weather radar system is shown in Fig 9. The equipment consists of a parabolic reflector, a transmitter-receiver with associated power supplies, and an indicator together with a control panel. Semiconductor

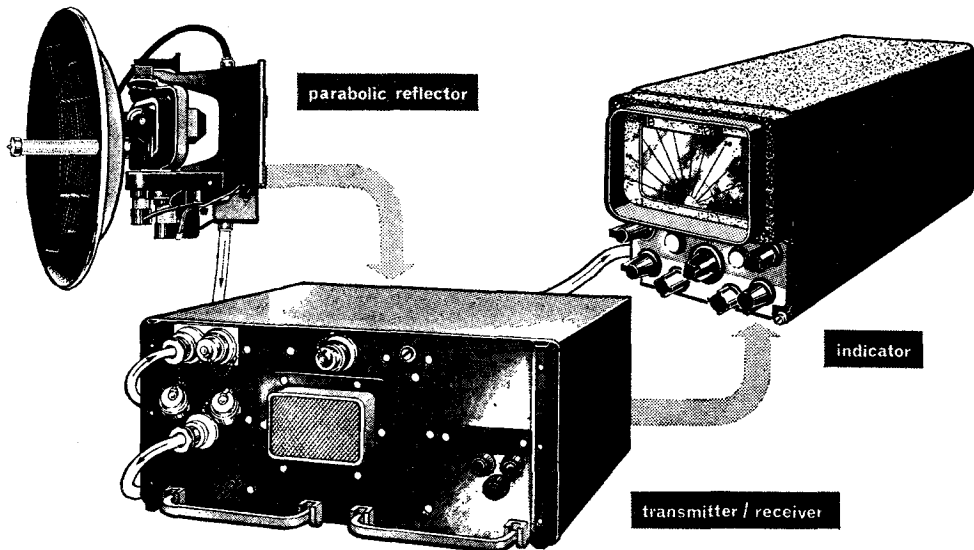


FIG 9. COMPONENTS IN A TYPICAL AIRBORNE WEATHER RADAR SYSTEM

devices are used extensively (exceptions include the transmitter magnetron and the display c.r.t.) and the reliability of the equipment is high. The characteristics of this equipment are shown in Table 1.

Frequency	9.4GHz (X band)
Peak Power	15kW
Pulse Duration	2.2 μ s
PRF	400 p.p.s.
Receiver Noise Factor	12 db
Aerial Beamwidth	3.5°
Maximum Range	150 n.m.

TABLE 1. CHARACTERISTICS OF TYPICAL WEATHER RADAR EQUIPMENT

Display. Airborne weather radars use an off-centred p.p.i. display on which electronic range marker rings are also displayed. A graticule, on which bearing lines are engraved, is placed over the front of the c.r.t. (Fig 10a).

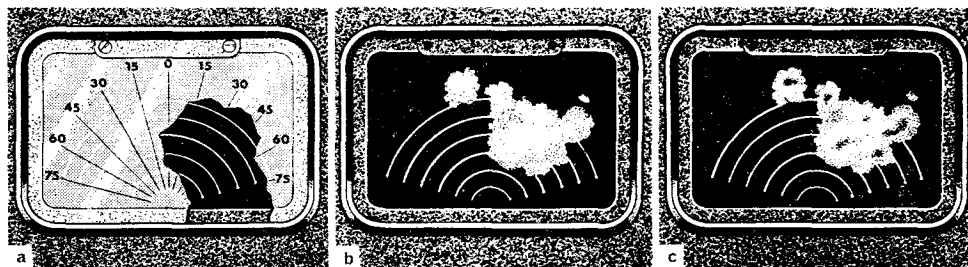


FIG 10. WEATHER RADAR PPI DISPLAYS

Many displays make use of 'iso-echo contour techniques' for pin-pointing areas of high turbulence. With this system, echoes above a pre-determined amplitude are *inverted* so that a black hole appears in the bright paint to indicate areas of heavy precipitation. The *width* of the paint surrounding the black hole shows how quickly the precipitation rate is changing from zero value at the edge of the cloud to the high rate existing in the hole. Rapid changes in precipitation rate, i.e. narrow white areas around the hole, indicate severe turbulence. Such areas must always be avoided by aircraft.

Fig 10b illustrates the display of a typical thunderstorm cloud area, without iso-echo contour techniques. Fig 10c is the same display with iso-echo contour applied.

A 'tilt' switch is usually fitted, operation of which tilts the scanner from its horizontal position within the limits of $\pm 15^\circ$. This enables two additional functions to be carried out:

a. Cloud height determination. It is advisable for an aircraft to avoid areas of high turbulence. This may be done by flying round the thunderstorm clouds or, if the frontal area is extensive, by flying above or below them. In the latter case it is necessary to determine the *height* of the cloud. The range of the cloud is first obtained from the p.p.i.; the scanner is

then tilted slowly to the position where the cloud disappears from the display; knowing the range and the tilt angle the height of the cloud may then be calculated and the necessary avoiding action taken.

b. Ground mapping. By tilting the scanner down by an appropriate angle (related to aircraft height) radar returns from the ground will produce a map on the p.p.i. The map obtained may be used as a valuable aid to the navigation of the aircraft.

Collision-avoidance Systems

One of the hazards that an aircraft has to face is the danger of air-to-air collision. With large numbers of aircraft operating in crowded air-spaces, and with the continued increase in air traffic, this hazard is becoming progressively greater. To a large extent, the responsibility for avoiding collisions between aircraft in flight rests with the air traffic control system, and the aircraft separation imposed by controllers has been successful in reducing the number of incidents to a very low figure. However, it is now clear that some form of *airborne* collision-avoidance system is needed to remove some of the increasingly heavy burden borne by air traffic controllers.

Much work has been devoted to the development of a system suitable for installation in aircraft. To date, no airborne collision-avoidance system has been found acceptable, but many manufacturers continue to work on the problem. Many ideas have been proposed but the one that is rapidly being accepted for investigation by most firms is described in the following paragraphs.

It appears that a collision-avoidance system will involve the measurement of range, the rate of change of range (range rate), and altitude. With such a system each aircraft carries a computer which accurately controls the transmission time of the equipment, the allocated time of transmission being different for each aircraft. The air traffic control supplies the transmit time of all aircraft in an area to each computer. All aircraft transmit their *altitude* on a common frequency at the *specified time*, and the *difference* in microseconds between the allocated time of transmission and the actual time of reception is used to determine the *distance* to the transmitting aircraft.

At the receiving aircraft the incoming signal also experiences a Doppler shift which has a value proportional to the closing speed of the two aircraft. By measuring the Doppler shift the *range rate* may be determined.

The computer then divides *distance* by *range rate* to predict *time* of possible collision, advises on the necessary avoiding action based on the *altitude* information, and indicates when the danger is past.

Prototype equipment using this principle has been developed and the results obtained have been described as excellent. However, great precision in both timing and frequency stability is necessary and this poses problems that have not yet been fully solved. In addition, the system is a cooperative one and relies on all aircraft being fitted with the same equipment. This is an obvious limitation. Nevertheless, since some form of collision-avoidance system is essential, it is clear that the development work will continue until the problem is solved.

CHAPTER 4

PRIMARY GROUND RADAR DEVELOPMENTS

Introduction

In this chapter we wish to consider some of the advances that have been made in primary ground radar systems.

A modern ground radar search system must be capable of detecting a target at very long ranges and be able to define that target in terms of range, Doppler velocity, and angular direction with very great accuracy. This task has two distinct stages:

- a. The radar must be able to *detect* a signal under heavy noise or clutter conditions.
- b. The radar must be able to *extract* the necessary information about the target from the detected signal with a known degree of accuracy.

Noise is the limiting factor in the first of these processes. It has been stated several times in this book that noise is the main factor which ultimately limits the usefulness of any radar system. Great efforts have been made in the past to reduce the amount of noise generated within the equipment itself. Low-noise components have been developed; low-noise r.f. amplifiers, such as the t.w.t., the parametric amplifier, and the maser, are being used increasingly in radar receivers; mixers with improved noise factors are being introduced; noise-cancelling networks have been built into radar transmitters. These steps are quite apart from those that have been taken to reduce *clutter*—circular polarization, m.t.i. systems, swept gain circuits, logarithmic receivers, and so on. Despite all these considerable advances, efforts to improve the received signal-to-noise ratio continue. Much of the development in both the transmitting and receiving sections of the ground radar equipment has this aim in view.

The requirements of high accuracy in range and Doppler velocity measurements have led to new developments in radar waveform generation. This task has been eased with the introduction of microwave amplifier radars (see p.569). The first reliable source of microwave power was the magnetron, which produces *randomly-phased* pulses. The more-recently developed amplifier valves, such as the klystron, the t.w.t., and the amplitron, enable *coherent* radar waveforms to be generated at low power and then amplified for radiation from the aerial. Coherent systems are those in which the carrier from pulse to pulse is gated from the same continuous wave in such a way that the *phase* of the radiated energy is the same for each pulse. Waveforms generated in this way enable radar measurements to be made with improved accuracy.

In this chapter we shall first consider some of the ways in which the problem of detecting radar signals in noise has been tackled. We shall then consider some of the factors affecting the accuracy of radar measurements and the steps taken to improve this accuracy.

Correlation Processes

Where it is necessary to detect, and extract information from, signals that have a very poor signal-to-noise ratio, the signals can be dealt with satisfactorily only by using *correlation* techniques. Correlation is a function that gives a measure of the *relationship* between two inputs, i.e. it shows how alike they are. For example, an operator viewing a c.r.t. display is performing correlation in recognizing a target. He has in his mind's eye an image of the display that corresponds to the target he wishes to find on the c.r.t. If a reasonably similar image appears it may be recognized by the operator as corresponding to the required target. In receivers employing correlation techniques this function is performed automatically by inserting suitable electronic circuits.

There are two correlation processes: *auto-correlation* and *cross-correlation*. In the first, the input signal is correlated with a replica of itself; in the second, a locally-generated 'matched' reference is used for correlation with the input signal. Auto-correlation is not much used in radar

receivers because if the input signal is noisy so also is the reference (the signal itself), whereas in a cross-correlation receiver the reference signal can be made noise-free.

Cross-correlation Receiver

Fig. 1 illustrates in outline one form of cross-correlation receiver. The transmitter radiates a signal V_1 at time t . The input to the receiver is then the signal V_1 at time $(t + T_0)$, together with noise, T_0 being the echo delay time. This input is applied to a comparator circuit along with the reference signal. The reference signal is the delayed replica of the transmitted pulse and is

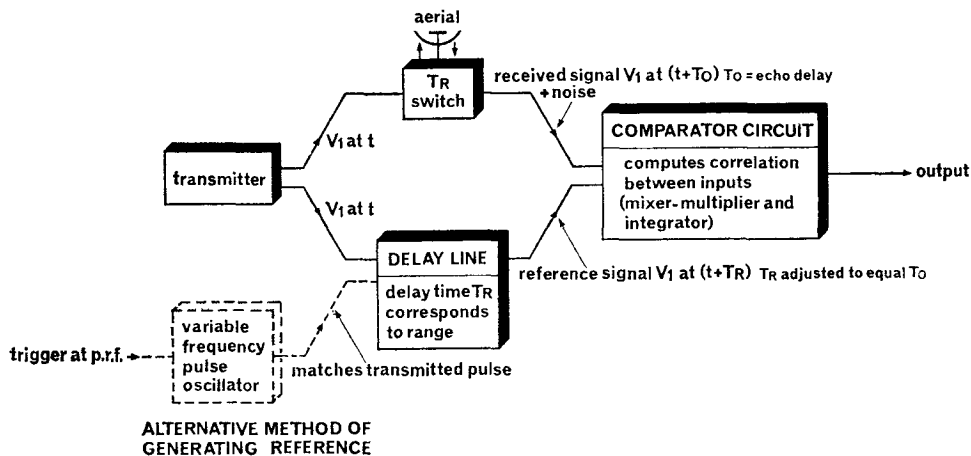


FIG. 1 EXAMPLE OF CROSS-CORRELATION RECEIVER, OUTLINE ONLY

available at the output of the delay line as a signal V_1 at time $(t + T_R)$. The delay T_R introduced by the delay line is adjusted to match the echo delay time T_0 and is proportional to the *range* to the target. The comparator computes the correlation between the received signal plus noise and the reference signal and produces an output only when there is the necessary relationship between the two inputs in terms of time, frequency, and pulse duration. The circuit arrangement of the comparator is such that the two inputs are multiplied in a mixer and the average value of the product from an integrator circuit provides the output. Only the signal component of the input is correlated with the reference and noise is practically eliminated.

With a fixed delay line of delay time T_R , the circuit of Fig 1 can test for the presence of a target *at only one range*. To check for targets at other ranges, the delay time has to be varied. It is usually more convenient to provide additional parallel channels, where each channel contains a delay line with a fixed delay time corresponding to a particular range.

If the frequency of the echo signal is not known exactly (e.g. if it has been Doppler-shifted) several correlation receivers must be used in parallel to cover the whole band of frequencies in which the echo signal may be expected. The frequency of the reference signal must be adjusted accordingly for each receiver. Under these conditions the reference signal may be obtained from a triggered pulse oscillator that produces pulses of the required frequency, p.r.f. and pulse duration, delayed by a suitable delay line. This alternative arrangement is also shown in Fig. 1.

Matched-filter Receiver

The matched-filter receiver, mentioned briefly in p.405, is a system that provides the maximum available signal-to-noise ratio under given conditions. In such a receiver, the i.f. stage is designed

as a filter whose frequency-response function is such that it passes the desired signal whilst rejecting much of the noise.

We noted in p.405 that the bandwidth of the filter must be carefully 'matched' to the spectrum of the desired signal so that neither too much noise nor too little signal is received. If the bandwidth is too large, excess noise is received; if it is too small, some of the signal energy is lost. Thus, there is an *optimum bandwidth* at which the signal-to-noise ratio is a maximum. The matched filter is designed to provide this bandwidth.

In radar, a signal is received as an echo from a target and, during reception, noise is added to the signal. Thus, we have a certain signal-to-noise ratio at the input to the receiver. The normal radar receiver may be considered to act as a type of filter, the 'filter' being in that part of the receiver where the frequency-response function is determined, namely the i.f. stage. The output then consists of a filtered signal and filtered noise, having a certain signal-to-noise ratio. The optimum receiver-filter is that which accentuates the signal component and attenuates the noise component to provide the *maximum possible* output signal-to-noise ratio for a given input ratio. Maximum output signal-to-noise ratio also gives maximum probability of detection so that this type of receiver provides the most reliable detection. For a receiver-filter to provide the maximum possible output signal-to-noise ratio, it must have a frequency-response function which matches the frequency spectrum of the target echo. Such an arrangement is called a *matched filter*.

It can be shown mathematically that the output of the matched filter is the same as would be obtained by cross-correlating the input signal with a replica of the transmitted signal. Because of this, the matched-filter receiver is often said to be *equivalent* to the cross-correlation receiver. Both give similar results although the circuits necessary to implement each are different. For some applications the matched-filter receiver is easier to design and build; for others, the cross-correlation receiver is easier. For the matched filter the replica of the transmitted signal is 'built-in' to the filter in the form of its frequency-response function. Since the latter is matched to the frequency spectrum of the signal, the filter provides the maximum possible signal-to-noise ratio and the maximum probability of detection.

In practice, the matched filter cannot always be obtained exactly. However, a close approximation to the matched filter for a rectangular pulse can be provided by correct adjustment of relatively simple circuits. We have already seen that the optimum arrangement consists of an i.f. bandpass filter whose frequency-response function has the same shape as the echo pulse spectrum envelope. For a rectangular pulse, a good approximation to this occurs when the i.f. bandwidth is adjusted to a value of $\frac{1}{p}$, where p is the pulse duration. More complex arrangements can be provided to match the pulse spectrum to a greater degree of accuracy, but these will not be considered here.

Threshold of Detection

Having produced the best available signal-to-noise ratio by the use of matched-filter or correlation techniques in the early part of the receiver, the next stage to consider is the detector itself.

The problem in detecting weak signals in noise is to determine whether the output from the receiver is due to noise alone or to signal-plus-noise. A radar operator viewing a c.r.t. display makes this decision on the basis of what he sees on the screen. However, when detection is being carried out by electronic means without the aid of an operator, a decision on a definite 'yes-no' basis is needed.

In practice, most radar detection decisions are based upon determining whether the receiver output is greater or less than some *predetermined threshold level*. If the output exceeds the threshold, a signal is assumed to be present; if the output is below this level, it is assumed to be

due to noise alone. The purpose of the threshold level is to divide the output into regions of 'detection' and 'no detection' (Fig 2).

It is possible with this method to mistake noise for a signal when, in fact, noise alone is present. This happens when the noise is great enough to exceed the threshold level, and a 'false alarm' results. On the other hand, it is possible for an input to be wrongly considered as noise when, in fact, the signal is present. This happens when the signal is too weak to lift it over the set threshold level, and a 'missed detection' results. The threshold level must be set to give a compromise between these two types of error. A large threshold reduces the number of false alarms but increases the number of missed detections. The opposite is true for a small threshold. The actual setting of the threshold level depends upon the operational function of the radar, i.e. whether it is better to accept false alarms or to avoid missed detections.

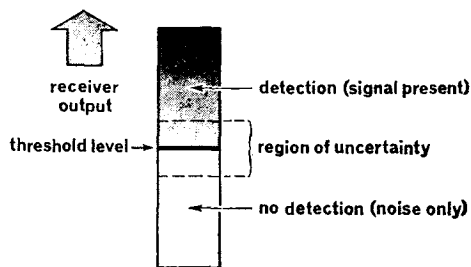


FIG. 2 THRESHOLD LEVEL IN DETECTION

Types of Detector

There are three main types of detector that may be used in a radar receiver:

a. *The envelope detector.* This is the familiar diode detector considered elsewhere in this book (see p.418). It depends for its operation on the *amplitude* of the carrier envelope and makes no use of the phase information contained in the signal.

b. *The cycle-counting detector.* This type of detector (sometimes referred to as a 'zero-crossings detector') depends for its operation on the *phase* information contained in the signal. Amplitude information is lost.

c. *The coherent detector.* This type of detector makes use of *both* the phase and the amplitude information contained in the input signal and performs more efficiently than either of the other two types.

The envelope detector. Most present-day radars use this type of detector. It consists of a rectifying element and a low-pass filter to pass the modulation (video) frequencies whilst removing the carrier. The detector operates either as a linear detector or as a square-law detector. For small signal-to-noise ratios the square-law detector gives the better results, whilst for large signal-to-noise ratios the linear detector is better. But the practical differences are small.

The logarithmic detector (see p.432) is also an envelope detector. It is used where a large dynamic range is needed to prevent receiver saturation due to clutter.

The cycle-counting detector. This type of detector effectively *counts* the number of times the input waveform crosses the zero-voltage axis in a given time. The number of crossings depends upon the frequency of the waveform and it also depends upon whether signal alone is present or signal-plus-noise. In Fig. 3a, where signal alone is assumed to be present, six zero-crossings are counted in the time T. In Fig. 3b, the number of crossings has increased to nine, the increase being due to noise.

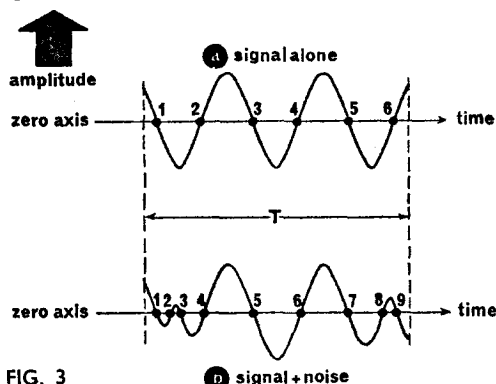


FIG. 3 OPERATION OF CYCLE-COUNTING DETECTOR

Thus, when noise is present, the number of zero crossings of an input signal in the given time increases. In general, the *smaller* the signal-to-noise ratio, the *greater* is the average number of crossings. A target is detected when the average number of zero crossings in a given time is *less* than a set threshold level. A frequency-counting meter may be used to count the zero crossings.

The results obtained with the envelope

detector and the cycle-counting detector are similar. The cycle-counting type has an advantage in that variations in receiver gain have no effect on its operation. This is not so for the envelope detector where variations in gain affect the *amplitude* of the signal and hence detector output. The main disadvantage of the cycle counting detector is that more circuits are needed than with the envelope detector.

The coherent detector. Fig. 4 illustrates the outline of a coherent detector. There are two inputs to the balanced mixer:

- a. The i.f. signal of known frequency and phase.
- b. An input from the reference oscillator, assumed to have the *same* frequency and phase as the i.f. input.

The output of the balanced mixer is applied to a low-pass filter which allows only the d.c. and low-frequency modulation components to pass whilst rejecting the higher frequency components. By mixing two inputs of equal frequency and phase, the coherent detector effectively *translates* the carrier frequency to d.c. It does not 'extract' the modulation envelope as does the envelope detector. In a normal balanced mixer (see p.306) the d.c. component is thrown away and the i.f. is retained; in Fig. 4, the d.c. is retained and the i.f. discarded. The coherent detector utilizes both amplitude and phase components of the input signal and is, therefore, more efficient than either of the other two types of detector, especially for the retrieval of signals buried in noise.

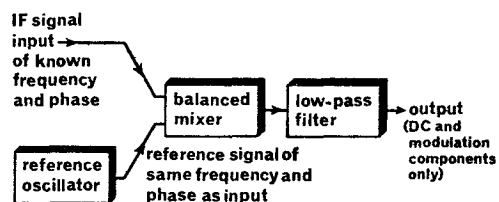


FIG. 4 OUTLINE OF COHERENT DETECTOR

The coherent detector is similar in some respects to the phase-sensitive detector used in m.t.i. radars (see pp.450 and 501). The main difference is that the reference signal in the phase-sensitive detector need not be of the same *phase* as the input signal, whereas in the coherent detector this is a requirement.

Integration of Echo Pulses

When a target is being scanned by a pulsed radar, several target echoes are received back at the radar during the time that the aerial beam is sweeping past the target. The number of echoes received depends upon the radar p.r.f., the aerial beamwidth, the scanning rate, and the range to the target. Since the probability of detection depends upon the signal-to-noise ratio it is necessary for the radar to sum or '*integrate*' all the pulses received during the time that the aerial beam is on the target to provide a large signal energy. In a conventional p.p.i. display all the echoes from a target are integrated by the display itself to form a single well-painted blip and using this blip the radar operator makes his estimate about the target.

In some radar applications, the integration of radar pulses must be done automatically without the aid of an operator viewing a display. The integration may be carried out either *before* or *after* detection depending upon the type of receiver being used. In non-coherent detection (e.g. envelope detection) no attempt is made to correlate the phases of transmitted and received signals. The signal returns from each target must be detected *separately* and added together in the video stages of the receiver *after detection*. In coherent detection, the phase relationship between transmitted and received signals is maintained. This makes it possible to add successive pulses during a scan *before detection* to obtain a direct enhancement of the signal-to-noise ratio. The improvement in signal-to-noise ratio is greater for pre-detection integration than for post-detection integration, so that longer detection ranges are possible especially at very low values of input signal-to-noise ratio.

Although pre-detection integration is the more efficient system it is more difficult to implement than integration in the video stages. Because of this, *video integration* is more common.

We shall consider two methods of video integration: the tapped delay-line integrator, and the binary (digital) integrator.

Tapped delay-line integrator. The outline of a typical system, providing a relatively simple method of integration, is shown in Fig 5. The time delay through the line is made equal to the *total* integration time which, in turn corresponds to the length of time that the radar beam illuminates a typical target. The taps are spaced at intervals corresponding to the pulse repetition period and the number of taps equals the average number of pulses that are expected to be integrated during the target illumination time. The outputs from the taps are tied to a common line to provide an output equal to the *sum* of all the pulses. By integrating the pulses in this way the signal energy is concentrated to provide a large signal-to-noise ratio.

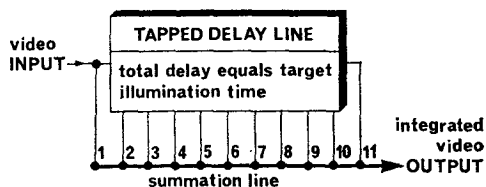


FIG. 5 TAPPED DELAY-LINE INTEGRATOR

Binary integrator. In binary integration the number of pulses received in a given time is compared with the number to be expected from a typical target during each scan. If the number received exceeds a set threshold level, a target is assumed to be present and an output pulse is generated. Thus, if nine pulses per scan are expected, the threshold may be set at a level of seven pulses per scan; then any number received in excess of seven indicates the presence of a target. A block diagram and waveforms illustrating this technique are shown in Fig. 6.

The radar video is passed through a negative limiter so that only those signals whose amplitude exceeds the pre-set limiting level are allowed to pass to the output (waveform 2). The output of the limiter is applied to a pulse generator which generates a standard pulse (binary digit 1) if the video waveform exceeds the limiting value, and nothing (binary digit 0) if it does not. This output is shown by waveform 3.

For each aerial scan we shall have 'clusters' of pulses from targets at different ranges. The range gates select only those pulses corresponding to their appropriate time (range) intervals. Thus, if there happens to be a target within range interval 1, range gate 1 selects the binary digits 1 or 0 appearing during that time interval. The selected pulses are passed by the appropriate range gate to a binary counter. If the number of binary digits 1 counted during the selected time interval exceeds the set threshold level the count sampler is triggered and generates a pulse indicating that a target has been detected at a certain range interval. This process applies to all positions of the aerial so that during one complete sweep many target pulses at different ranges and bearings are generated and applied to the radar display or data computer.

Extraction of Radar Information

So far we have been concerned with target detection and we have discussed in general terms the modern methods of improving the probability of detection. However, once the target has been detected and a target pulse produced, it is necessary to obtain information about the target from that pulse. The information that may be extracted from a target echo includes:

- a. Range to the target.
- b. Doppler velocity of target.
- c. Angle of target in azimuth and elevation.

From these three items of information it is also possible to compute the present and future positions of the target. An outline of most of the methods used for providing these items of information is described earlier in this book and will not be repeated here.

Factors Affecting Accuracy of Radar Measurements

We know that the ability of a radar system to detect the presence of a target is limited by noise. Noise is also one of the factors that limits the *accuracy* of radar measurements and, for

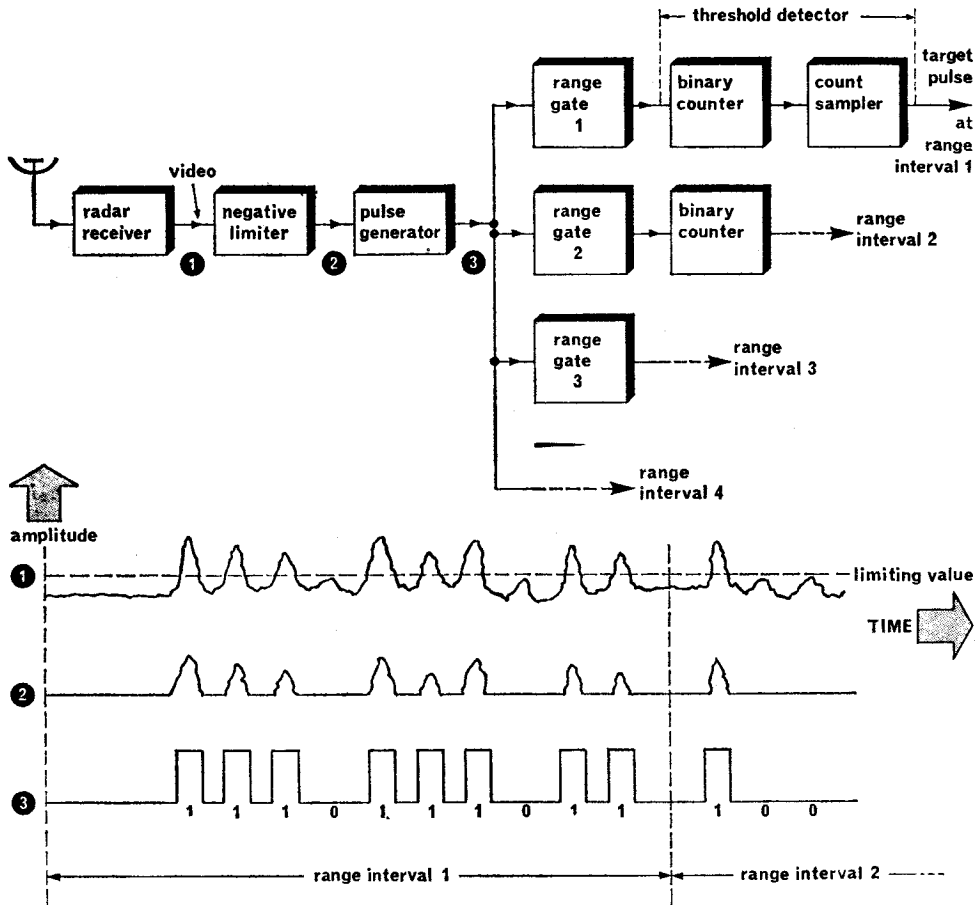


FIG. 6 BINARY INTEGRATION, BASIC IDEA

accurate measurements, large signal-to-noise ratios are necessary.

For pulsed radar we require a pulse with a steep leading edge to provide high accuracy in *range* measurement. To produce such a pulse, a signal with a *large bandwidth* is needed. The radar waveform that gives the most accurate range measurement is the one with the largest effective bandwidth.

Similar considerations apply to accurate measurement of Doppler-velocity and angular direction. But in addition, to avoid *ambiguity* in radar measurements, we require a signal that is *long in duration* (see p.460).

Thus, we can say that to obtain high accuracy in radar measurements of range, Doppler velocity, and angular direction, without ambiguity, we require:

- a. A high signal-to-noise ratio.
- b. A signal of wide bandwidth.
- c. A signal that is long in duration.

In fact, we can go further and state that the factor which has the greatest effect on detection probability, measurement accuracy, and ambiguity is the *bandwidth-pulse duration product* of the radiated signal. This factor is often referred to as the *Bp factor*, where B is the bandwidth and p the pulse duration. For good detection probability, high accuracy, and no ambiguity, the Bp factor should be large.

We have stated earlier in this book that a short duration pulse provides a wide bandwidth (essential for accuracy) whilst a narrow bandwidth is associated with a long duration pulse (necessary to avoid ambiguity). It would therefore appear that it is impossible to obtain a long duration signal and a wide bandwidth simultaneously, i.e. that there is no way in which the Bp factor can be made large. However, there are methods by which this can be done (e.g. pulse compression techniques) and these are considered later.

Transmitted Waveform

The design of the transmitted waveform depends upon what is required from the radar in terms of detection probability, measurement accuracy, ambiguity, and resolution. If the receiver is designed as a matched filter for the particular transmitted waveform, the probability of detection depends only upon the signal-to-noise ratio and is not affected by the waveform. In other words, so far as the transmitter is concerned, the *total energy* to be radiated is the factor that affects detection probability. On the other hand, the accuracy, ambiguity, and resolution requirements of the radar depend to a large extent upon the *design* of the transmitted waveform.

Accuracy. If the receiver is designed as a matched filter or as a correlation receiver, its output is proportional to the degree of correlation between the received signal and the stored replica of the transmitted signal. When the received signal 'matches' the stored replica exactly, the receiver output is a maximum. Thus, when a target is present at a time interval corresponding to a target delay time of T_0 , the 'correlation function' peaks to a maximum when T_0 equals T_R , T_R being the stored replica delay time. Ideally, the output should be a maximum under these conditions, and zero elsewhere. In practice, this ideal condition is not possible; there is some output from the receiver at other values of T_R in the vicinity of T_0 (Fig. 7). The extent of the spread about this peak determines the *accuracy* with which T_0 , and hence target range, may be measured.

Ambiguity. An ambiguous measurement is one in which two or more readings give similar results, but only one of the readings is the true one; the uncertainty gives rise to ambiguity. Examples of ambiguities are second-time-round echoes with pulsed radar, and blind speeds with m.t.i. radar. If the correlation function of Fig. 7 has only *one peak* there is no ambiguity; if it peaks for values of T_R other than the true target delay time of T_0 , ambiguity arises.

Resolution. This is the property of the radar to distinguish between targets. Resolution is needed when similar targets are close together, as with several aircraft flying in formation. Resolution is also needed to select a wanted target in the presence of clutter. For good resolution the correlation function of Fig 7 should show a single very narrow peak in the vicinity of T_0 . The radar can then discriminate between two targets close together in range.

The actual shape of the correlation function diagram of Fig. 7 is determined by the *waveform* of the radar transmitted signal. Different waveform shapes, pulse durations, and so on produce different correlation function diagrams. The correlation function can be described mathematically, worked out for a given radar system, and drawn as a Fig. 7 graph. This will not be done here because the mathematics are too difficult for this book. In fact, it would be better if the method could be made to work the other way round, i.e. decide on the correlation function diagram required and design the radar waveform to provide this. It has not yet been possible to do this. Ideally as we have seen, the waveform should be such that it produces a single narrow peak in the correlation function at T_0 equals T_R . A waveform that produces such a function provides accurate radar measurements in range, with good resolution and with no ambiguity. A pulse with a *large Bp factor* approaches this ideal.

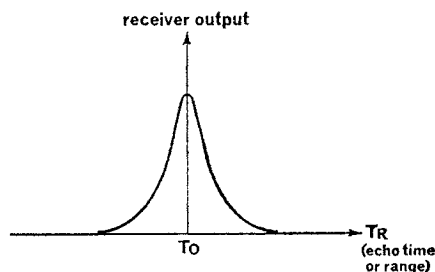


FIG. 7 RANGE CORRELATION FUNCTION

So far we have been concerned with accuracy and resolution in *range*. Similar considerations apply to the ability of a particular waveform to resolve targets with *different Doppler velocities* but at the same range. We could construct a graph similar to that of Fig. 7 with a Doppler velocity axis in place of the T_R (range) axis. This graph would then give the correlation function for a given waveform in terms of Doppler velocity. In practice it is easier to combine the correlation functions for Doppler velocity and range on a single 'ambiguity diagram'.

Ambiguity diagram. The effect of the transmitted waveform on both the Doppler velocity (frequency shift) and range (time delay) may be illustrated by an extension of the correlation function of the received signal. A three-axes graph results, with receiver output plotted against echo time (range) on one set of axes, and against frequency shift (Doppler velocity) on the other set of axes, at right angles to the first set (Fig. 8). This graph is used to indicate the accuracy, resolution, and ambiguity provided by a given transmitted waveform in *both range and Doppler velocity*. A matched-filter receiver is assumed.

The ambiguity diagram produced by a waveform approaching the ideal is illustrated in Fig. 8b. The peak of the graph cannot exceed a given height (determined by the energy in the transmitted signal), and the graph encloses a fixed volume. The waveform producing the diagram of Fig. 8 does not cause ambiguities in radar measurements, because the function has a single peak. But this peak may be too broad to give the required accuracy and resolution in range, in Doppler velocity, or in both. The transmitted waveform may be adjusted to provide a narrower peak on the ambiguity diagram but, since the fixed volume under the graph has to be maintained, the function has to be raised elsewhere and this may produce additional peaks, giving rise to ambiguity. From any given transmitted waveform it may not be possible to satisfy the required accuracy and ambiguity simultaneously; a *compromise* is the usual result.

The transmitted waveform that comes nearest to producing the ideal ambiguity diagram consists of a single pulse with a large bandwidth-pulse duration (Bp) product. As mentioned earlier, a long duration pulse is usually associated with a narrow bandwidth. However, the bandwidth may be increased by applying phase or frequency modulation to the r.f. energy *within* the pulse. Modulation always produces sidebands, causing an increase in the overall bandwidth of the radiated signal. This is done in the pulse compression system discussed later.

Summary. A transmitted waveform satisfies the requirements of detection if its energy is sufficiently large. The ability of a particular waveform to satisfy the requirements of accuracy, resolution, and ambiguity can be determined by examining the ambiguity diagram produced by that waveform in a matched-filter or correlation receiver. For a transmitted waveform with a large Bp product, the ambiguity diagram shows that it is possible to determine simultaneously to a high degree of accuracy both the frequency shift (Doppler velocity) and the time delay (range) of a target.

Pulse Compression

In a pulsed radar system we are faced with the problem that for good range resolution and accuracy we normally require a *short pulse*, whilst to avoid ambiguity and to provide good

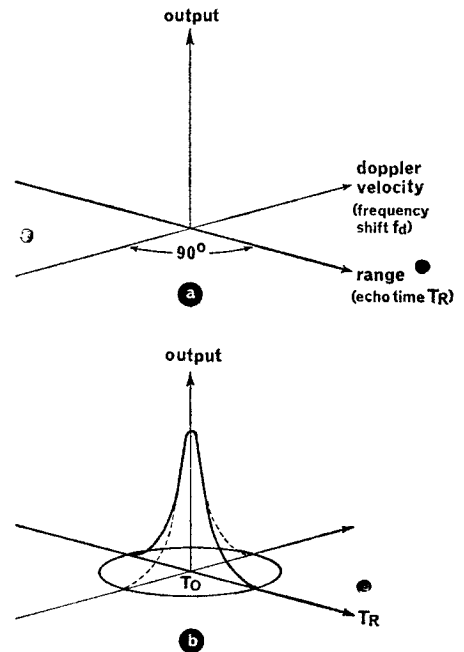


FIG. 8 AMBIGUITY DIAGRAM FOR A GIVEN TRANSMITTED WAVEFORM

probability of detection we usually need a *long pulse*. Pulse compression is a means of operating a radar with long pulses to obtain the resolution and accuracy of a short pulse but the detection capability of a long pulse. This appears to be a case both of 'having our cake and eating it'.

Pulse compression *stretches* a short low-power pulse into a long *modulated* pulse prior to amplification and transmission. The receiver is designed to act on the modulation to *compress* the pulse into a very much shorter one, comparable with the original short pulse at the transmitter. We saw earlier that good resolution and accuracy without ambiguity can be obtained with any pulse provided its B_p product is large enough. In pulse compression, the radiated waveform has a long duration and, by modulating within the pulse, its effective bandwidth can also be made large. Thus, the necessary condition is satisfied.

The energy contained in a radar pulse depends upon the peak power of the pulse and the length of time for which the pulse is transmitted (energy = power $\times$ time). Thus, to transmit a pulse containing sufficient energy to provide good probability of detection, the peak power must be large or the pulse duration long, or both (Fig. 9).

In many radars, the pulse duration is limited to small values to provide good range accuracy and resolution, so that the required energy per pulse must be obtained with a large peak power. In long-range radars it may not be possible to obtain a peak power large enough to provide adequate detection capability because of the danger of voltage breakdown in components. Such radars are said to be '*peak-power-limited*'. Conventional radars tend to be limited by peak power ratings long before optimum *mean* power levels are reached. In such cases the required energy per pulse can be obtained only by transmitting a longer pulse. In pulse compression radar the use of a long pulse means that much more energy can be provided in each pulse and this improves the probability of detection. The peak power rating for unpressurized waveguide at X-band is about 250 kW, and this applies whether the power is sustained for $0.1\mu\text{s}$ or for $10\mu\text{s}$. With the longer pulse, 100 times more energy per pulse is radiated at the peak power rating.

There are various methods by which pulse compression techniques may be applied to a radar. All the systems rely on 'stretching' an original pulse at the transmitter and modulating within the stretched pulse; at the receiver the reverse process occurs. We shall now consider methods that have been used in practice.

Phase-coded pulse compression. In this system of pulse compression the radar transmits a single long pulse which is built up of a number of consecutive short pulses, of equal duration and frequency but *differing in phase* in a known manner. This waveform may be generated by applying a single short initial pulse to a delay line which has, say, 13 taps spaced by the pulse duration of the short pulse (Fig. 10a). Phase-changers connected between these taps and a common output line are then used to generate a long phase-modulated pulse. In practice, the phase of the output from each of the taps is either unchanged (0°) or phase-changed by 180° (π radians) in a binary manner.

At the receiver a similar delay line, matching the long pulse duration, with 13 equally-spaced taps and 13 *reverse* phase-changers, causes the constituent short pulses to be brought into phase and time coherence to provide a *single short pulse* at the output (Fig 10b). Note that to bring the pulses together in *time* the delay applied to each pulse at the receiver is the *reverse* of that at the transmitter, i.e. the first constituent short pulse is delayed by the long pulse duration, whereas the last pulse in the sequence has no delay at the receiver. It is possible for the *same* delay line to be used both for transmission and reception, the input being applied to the opposite end and the phase-changers being switched to the reversed conditions for reception. It is important that the pulse compression filter at the receiver can be properly 'matched' to the constituent short pulses,

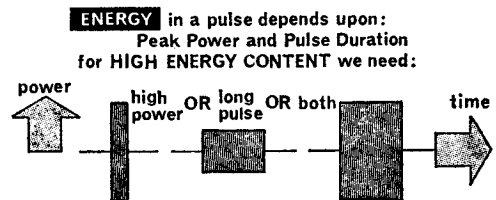


FIG. 9 WAYS OF OBTAINING HIGH ENERGY IN A PULSE

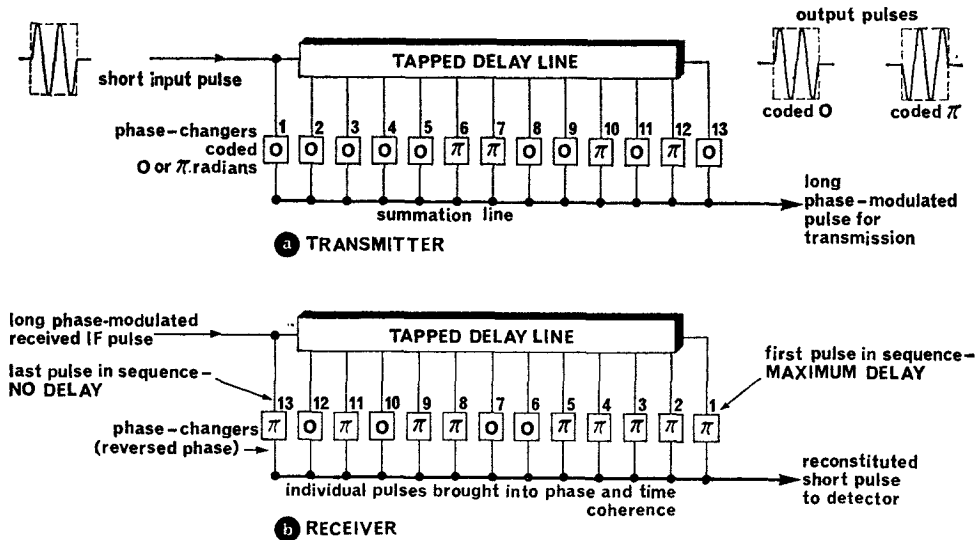


FIG. 10 PHASE-CODED PULSE COMPRESSION SYSTEM

otherwise timing and phasing will be upset and the filter will not properly compress the received long pulse. Range resolution and accuracy are determined by the duration of the *short* constituent pulses.

Fig. 11 shows an example of phase-coded pulse compression waveforms. The long transmitted pulse shown in Fig. 11a is made up of 13 short pulses of phase 00000 $\pi\pi$ 00 π 0 π 0. Fig. 11b shows the corresponding output from the pulse compression filter at the receiver.

The matched pulse compression filter at the receiver correlates the received waveform with the transmitted modulation. The short pulses within the pulse compression filter combine in such a way that the output contains additional small peaks either side of the main peak (Fig. 11b). A similarity may be seen between the illustration of Fig. 11b and the radiation pattern of an aerial, with its main lobe and sidelobes (see p.318). However, in Fig. 11b the lobes are in terms of time or *range* and are referred to as *range side-lobes*. They are usually small enough relative to the main lobe, corresponding to true target range, to avoid ambiguity. But they can create a resolution problem because the range sidelobes of a very strong signal may be greater than the main lobe of a weak signal. The range sidelobes may be reduced by passing the received signal, after compression, through a suitable filter.

If the stretched radiated pulse is subject to a Doppler shift, caused by the radial velocity of a target, then a different range lobe pattern is

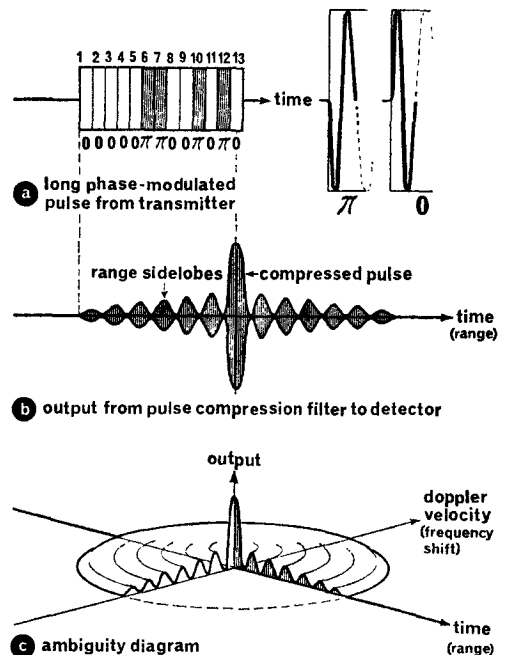


FIG. 11 PHASE-CODED PULSE COMPRESSION WAVEFORMS

produced. In fact, the full pattern in both range and Doppler velocity may be shown on an ambiguity diagram, similar to that of Fig. 8. The result may be as shown in Fig. 11c.

The 'coding' of a transmitted waveform in the way described above offers potentialities for the future. A radar waveform may be positively 'labelled' by means of coding and if the receiver is correctly matched to that particular waveform, only that signal will appear at the output of the receiver. The advantages of such a system are obvious, and investigation into the generation of binary-coded pulses continues.

Random frequency-modulated pulse compression. We know that a narrow pulse has a wide bandwidth so that the energy is spread over a wide frequency spectrum. If we apply a single narrow pulse to the system of Fig. 12, the pulse is shared between the four filters, each covering a

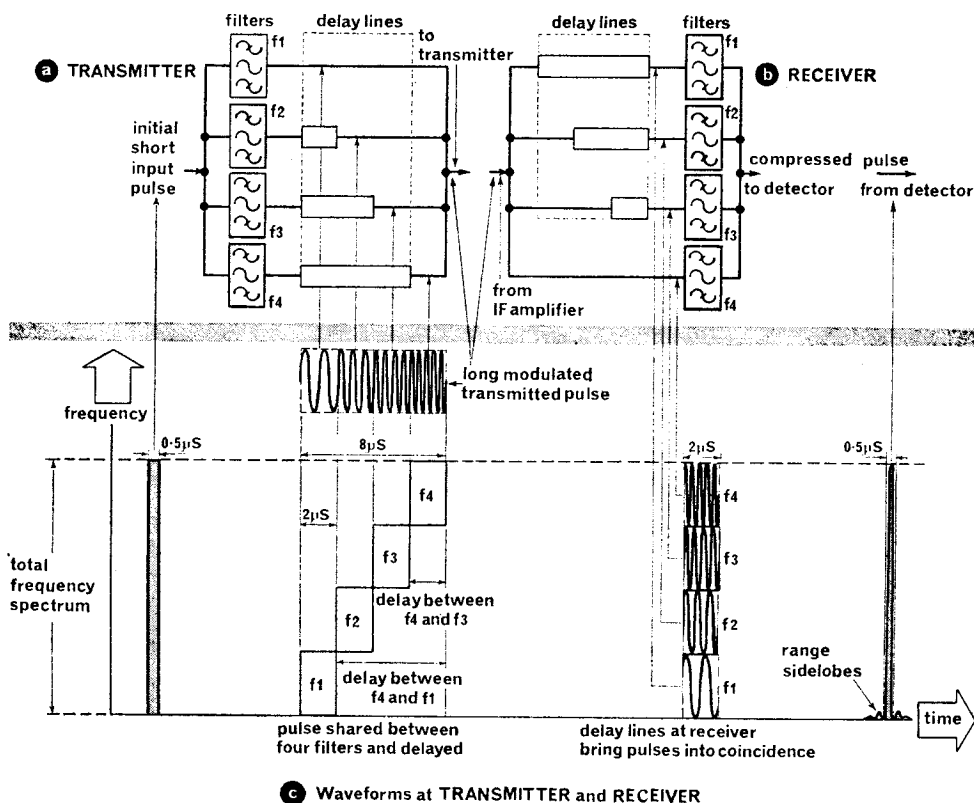


FIG. 12 OUTLINE OF RANDOM FREQUENCY-MODULATED PULSE COMPRESSION

different quarter of the total spectrum and tuned to the frequencies f_1, f_2, f_3 and f_4 . Each filter output occupies a *quarter* of the narrow-pulse frequency spectrum and, because of the reciprocal relationship between pulse duration and bandwidth, an output pulse of *four times* the duration of the original pulse is produced from each filter (Fig. 12c). With appropriate delay lines, the four pulses, each four times the original pulse duration, will generate a random frequency-modulated pulse of *16 times* the duration of the original pulse.

On reception, a pulse compression network with reversed delays (Fig. 12b) reconstitutes the short pulse to give the required range resolution and accuracy. In Fig. 12, the wide frequency

spectrum of the original narrow pulse has been maintained by modulating *within* the stretched pulse, and the pulse duration has been increased 16 times. The Bp product (sometimes referred to as the *pulse compression ratio*) has thus been increased and improved accuracy results.

Fig. 13 shows an example of random frequency-modulated pulse compression. In Fig. 13a the transmitted pulse is made up of four frequency-modulated pulses. Fig. 13b shows the corresponding output from the pulse compression filter at the receiver.

Linear frequency-modulated pulse compression.

We have seen how a stretched pulse may be built up from a number of consecutive short pulses, each of a slightly different frequency. By having *many* such frequency steps it is possible to reach a stage where there is a 'smooth' change of frequency during the long pulse. This is achieved in the *linear* frequency-modulated pulse compression system illustrated in Fig. 14.

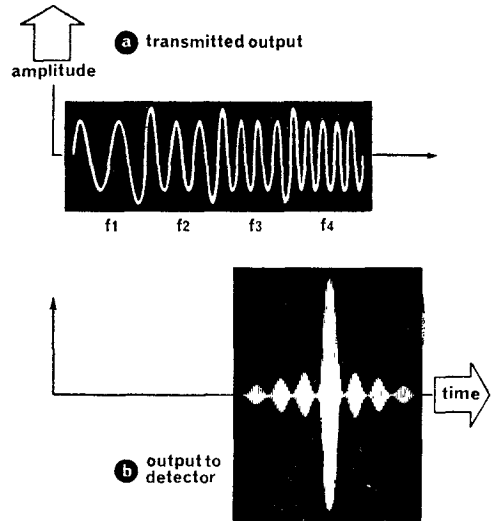


FIG. 13 RANDOM FREQUENCY-MODULATED PULSE COMPRESSION WAVEFORMS

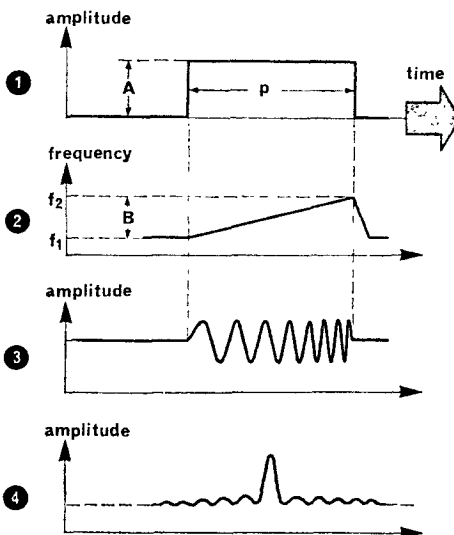
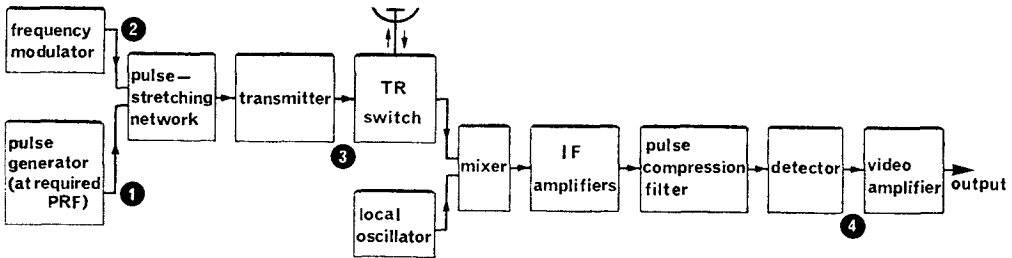


FIG. 14 LINEAR FREQUENCY-MODULATED PULSE COMPRESSION SYSTEM

With the exception of the stages concerned solely with stretching and modulating the pulse at the transmitter and compressing it at the receiver, the block diagram of Fig. 14 represents a conventional pulsed radar. The transmitted waveform consists of a pulse of constant amplitude A and duration p . The *frequency* of the pulse increases linearly from f_1 to f_2 over the duration of the pulse and the resulting transmitted waveform is indicated by waveform 3.

On reception, the f.m. echo signal is passed through the pulse compression filter, which is so designed that the *velocity* of propagation through it depends upon frequency. It may, for example, speed up the higher frequency components of the pulse relative to the lower frequency components. The result is that the energy contained in the original long-duration pulse is *compressed* into a much smaller duration pulse. The output from the detector consists of a large peak, corresponding to the true range or Doppler velocity, together with the usual range sidelobes (waveform 4).

By frequency-modulating within the pulse, the bandwidth has been increased and so has the bandwidth-pulse duration factor Bp , where B is the difference in frequency between f_1 and f_2 , and p is the original pulse duration. In some systems, the pulse compression ratio (the factor Bp) may be as high as 100. A system with a pulse compression ratio of 100 may be able to transmit a long duration pulse of, say, $10\mu\text{s}$ (frequency-modulated within the pulse) to provide a high energy level without excessive peak power. At the receiver this pulse is compressed to a duration of $0.1\mu\text{s}$, which is capable of providing the required high accuracy and resolution in both range and Doppler velocity without ambiguity. Still higher pulse compression ratios may be expected; ratios in excess of 1000 have been obtained experimentally.

Comparison of pulse compression and short pulse radars. Pulse compression is used when a long pulse is required to obtain a large transmitted pulse energy without danger of component breakdown, combined with good range resolution and accuracy, without ambiguity. With pulse compression, range resolution is usually poorer than that obtained with a short pulse because of the range sidelobes. It is also difficult to maintain the transmitter carrier frequency stable for the duration of a long pulse. On the other hand, in a peak-power-limited transmitter a short pulse may be incapable of providing adequate radiated energy for good probability of detection; and ambiguity in measurements also increases.

Frequency Agility in Ground Radars

Introduction. The term 'frequency agility' is used to describe a transmitter capable of operating at different frequencies in a given band on a pulse-to-pulse basis. Thus, the radiated frequency is slightly different for each pulse. This subject was mentioned briefly in p.578 when dealing with the dither-tuned magnetron. We now wish to look more closely at frequency-agile radars.

Frequency agility is applied to pulsed radar systems to improve radar range and bearing accuracy. It also has advantages over a fixed-frequency radar in reducing e.c.m. interference. Frequency agility may be applied to transmitters capable of being tuned over a given band of frequencies, but it is not essential for the transmitter to be tunable in this sense. The transmitter may be a magnetron-type, in which case a dither-tuned magnetron is used to provide frequency agility. Alternatively, the transmitter may be an m.o.-p.a. type, using a broadband amplifier whose output frequency is determined by a low-power, variable-frequency oscillator; the amplifier may be any of the microwave types considered in Chapter 2, i.e. multi-cavity klystron, amplitron, or t.w.t.

Limitations of fixed-frequency radars. The radar echo from a target is obtained from the energy scattered back towards the radar from the various parts of the target. The amplitude of the echo depends upon the phase relationship between the signals re-radiated from each 'separate' part of the target. This phase relationship is complex and depends, among other things, on the *frequency* of the incident energy and the *aspect* of the target (i.e. whether the target is flying along the radar beam, across it, and so on). For a fixed-frequency radar, if the phase relationships of the re-radiated signals from each part of the target are mutually destructive for one pulse, they are likely to be so for all pulses within a particular aerial scan. Thus, during this particular scan,

no target echo is received and the target is undetected. The converse is also true; if the phase relationships of the re-radiated signals from each separate part of the target are such that all the signals add up in phase, this is likely to be true for all pulses within that group; a strong echo is received in this case. We can say, therefore, that there is a *high correlation* between pulses within a given scan for a fixed-frequency radar, and the target returns are extremely variable from scan to scan, perhaps disappearing from the display for relatively long periods.

For a frequency-agile radar, each radiated pulse is at a different frequency. Hence, the phase relationships of the re-radiated signals from each part of the target are *different* for each pulse within a scan. Thus, although some of the echoes may be destroyed by wrong phase relationships, some strong echoes will remain and this enables the target to be detected *on every scan*. In other words, in frequency-agile radar, there is *very little correlation* between pulses within a given scan.

The required difference in frequency between pulses depends upon many factors, among them being the pulse duration, the p.r.f., and the target dimensions. In practice, the frequency shift between pulses is of the order of 1MHz to 5MHz.

Improvement of radar range. The performance of a search radar is normally measured by the probability of detecting a given target at a given range. If we plot a graph of detection probability against radar range for both fixed-frequency and frequency-agile radar, the result is as shown in Fig. 15. This graph shows that the frequency-agile system gives the better results, the detection probability being higher at greater ranges than for the fixed-frequency system for the reasons given earlier. For 100% probability of detection the range improvement is almost 2:1. Without frequency agility, this improvement in range could only be obtained with a *sixteen-fold* increase in transmitter power. Thus for long-range search radars, frequency agility offers distinct advantages over a fixed-frequency system.

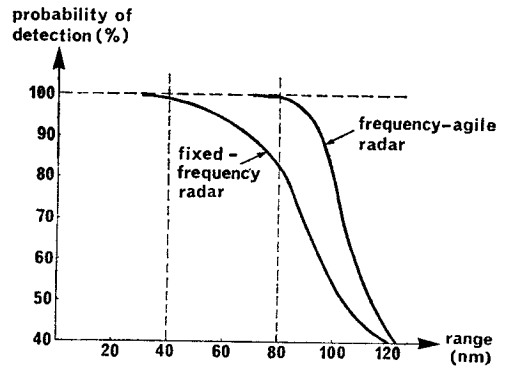


FIG. 15 COMPARISON OF FIXED-FREQUENCY AND FREQUENCY-AGILE RADARS

Reduction of clutter. With a fixed-frequency radar we have seen that there is a high correlation between pulses within a given scan. Thus, if phase relationships are favourable, it is possible for a very strong clutter signal to be produced during that scan. With frequency-agile radars there is very little correlation between pulses within each scan, because each pulse is at a different frequency, so that on average the clutter will remain much the same from scan to scan and will not 'peak' as in the fixed-frequency system.

Reduction of bearing angle errors. Target 'scintillation' has always been a problem in ground radar. A target is said to scintillate if the amplitude of the echo changes from pulse to pulse during the target illumination time. This may happen due to changes in target aspect relative to the aerial beam, or to changes in peak power output from pulse to pulse. One effect of target scintillation is to produce bearing errors.

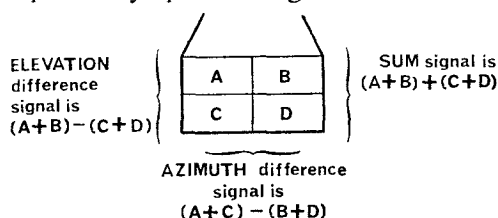
There will be very little error if the target does not scintillate during the target illumination time. There will also be little error if the target scintillates *many times* during the scan, because the returns will not be correlated so that their effects average out. If, however, the target scintillates only slightly during the scan so that the returns in a fixed-frequency system are partly correlated, bearing errors result. Thus, if the returns are weak while the aerial beam is on the target's left and strong while the beam is on the target's right, the displayed bearing will appear to be *shifted to the right*. The frequency-agile radar reduces this bearing error because returns during a scan are not correlated and scintillation effects average out. This is important in tracking radars.

Monopulse Tracking Radar

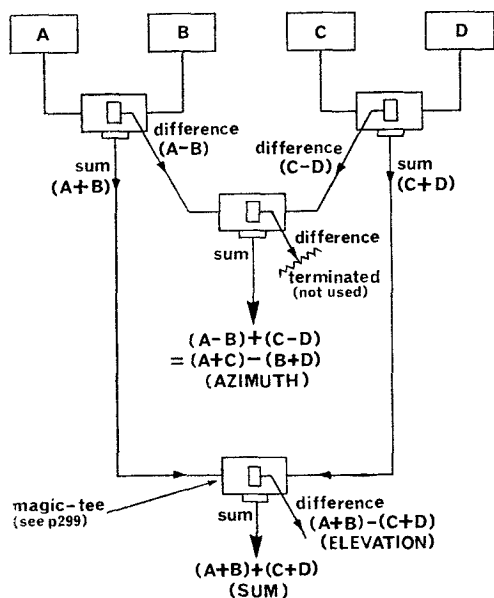
Introduction. Mention was made of the monopulse principle when discussing scanning methods in p.337. There we considered the application of monopulse to an airborne interception system. However, newer ground radars also use this technique to give accurate tracking of targets. It may, therefore, be helpful to have a closer look at this subject.

We noted earlier that conical scanning techniques used in tracking radars need at least four pulses during the scanning time to derive an error signal with which to position the aerial in azimuth and elevation. The use of four or more pulses during the scanning time has one big disadvantage: if the amplitude of the echo varies from pulse to pulse during this time because of scintillation effects, the tracking accuracy will be reduced. Pulse-to-pulse amplitude variations of the target echo have no effect on tracking accuracy if the angular measurement is made on the basis of *only one pulse*. Tracking radars that derive the angle-error information with only a single pulse are referred to, appropriately, as *monopulse radars*.

The monopulse principle. Monopulse radar is used to determine the angle to a distant target in both azimuth and elevation by comparing radar signals that are *simultaneously* received on several overlapping aerial patterns. Such a comparison relies on the fact that, although the pulse-to-pulse values of amplitude and phase may vary, the *relative* values within a single pulse depend only upon the angle of arrival of the signal.



a THE FOUR AERIAL FEEDS



b MONOPULSE BRIDGE, TYPICAL

FIG. 16 MONOPULSE AERIAL ARRANGEMENTS

To determine the angle error in both azimuth and elevation, *four* aerial radiation patterns are needed. The four beams may be generated with a single reflector illuminated by a cluster of four feeds (see p.337), although for some applications four separate aeriels have been used.

The four feed points (or the four aeriels) are connected to a monopulse bridge, which normally consists of an arrangement of four hybrid T-junctions (see p.299). The monopulse bridge senses the differences between the four signals and processes the information into a *sum* signal and *two difference* signals. The received sum signal provides the *range* measurement. The two difference signals are proportional to the angle of arrival of the energy and are used to derive an *error signal* which positions the aerial in azimuth and elevation. In practice, one difference channel determines the elevation angle whilst the other gives the azimuth angle.

Fig. 16a shows the schematic outline of the four feed points for the aerial, and Fig. 16b illustrates how these feeds are connected to the monopulse bridge. The sum channel is represented by the output $(A+B) + (C+D)$; the elevation difference channel is $(A+B) - (C+D)$; and the azimuth difference channel $(A+C) - (B+D)$.

The angle of arrival of the echo signal may be determined by measuring either the relative *amplitude* or the relative *phase* of the echo pulse received in each beam. Thus, we have a division of monopulse radars into amplitude-comparison

and phase-comparison types. It is also possible to use a combination of phase and amplitude measurements to provide the required results (see p.337). We shall confine our discussion to the more common amplitude-comparison type.

Amplitude-comparison monopulse. Let us consider the two overlapping beams used for determining the arrival of the signal in azimuth (Fig. 17a). A similar pattern applies for the elevation angle. Note that both beams are *offset* slightly from the common zero axis (*boresight*). The r.f. echo signals received from the two offset beams are combined to produce simultaneously two signals that are equal respectively to the sum and to the difference of the amplitudes of the two received signals. The *sum* of the two aerial patterns is shown in Fig 17b and the *difference* in Fig 17c. The sum pattern is used both for transmission and reception, whilst the difference pattern is used only for reception. The difference signal provides the magnitude and sign of the *error signal*. The sum signal, as well as providing the range measurement, is used as a reference in determining the *sign* of the error signal.

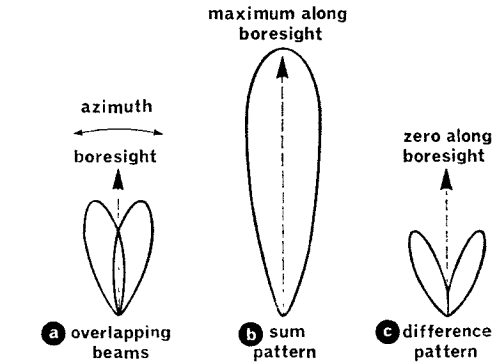


FIG. 17 EQUIVALENT SUM AND DIFFERENCE AERIAL RADIATION PATTERNS

A block diagram of an amplitude-comparison monopulse tracking radar is shown in Fig. 18. On reception, the outputs from the monopulse bridge at the sum and difference amplitudes are

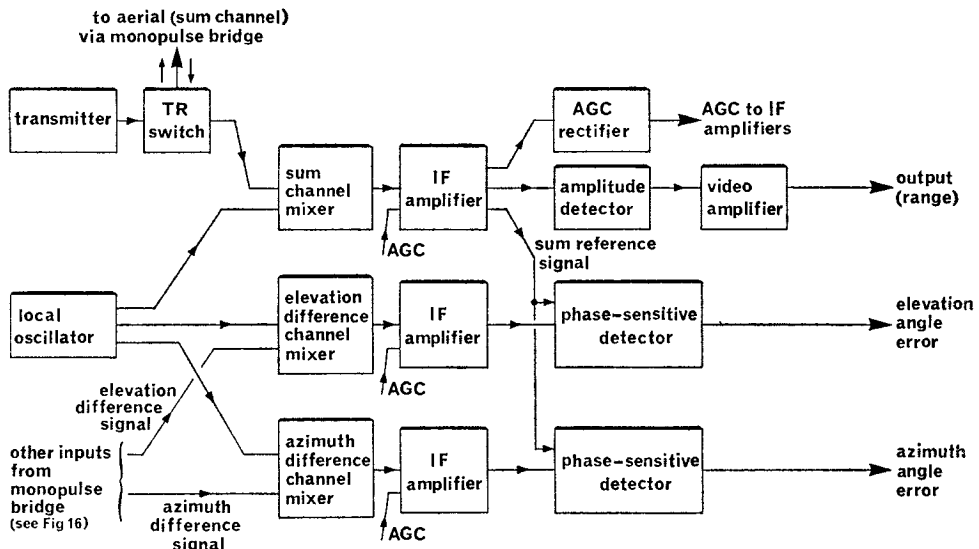


FIG. 18 BLOCK DIAGRAM OF TYPICAL AMPLITUDE-COMPARISON MONOPULSE RADAR

separately mixed with the output of a common local oscillator to produce the intermediate frequencies in the three channels. Note that the transmitter is connected via a TR switch to the *sum arm* so that radiation is effective from all four feed points.

The i.f. output of the sum channel is applied to an amplitude detector which extracts the video information to provide *range* to the target; it is also used to provide a reference to the phase-sensitive detectors to determine the sign of the error signal. The i.f. output of each difference channel is applied to separate phase-sensitive detectors, the other input to which is the reference signal from the sum channel. The output of the phase-sensitive detector in each difference channel

is then an error signal whose *magnitude* is proportional to the angular error of the target from boresight and whose *sign* is proportional to the direction of the angular error (see p.450 on phase-sensitive detectors). The angle error signals are normally used to operate a servo-control system to position the aerial in azimuth and elevation, the aerial being moved to the position where the error signal becomes zero; the target then lies along the boresight of the aerial. As the target position changes, so the aerial continues to track automatically.

It is worth noting that in amplitude-comparison monopulse the main purpose of the phase-sensitive detector is to provide the *sign* of the error signal. In fact, the i.f. output from each difference channel could be applied to an amplitude detector (as in the sum channel) to provide accurate angle error *magnitude*. The inputs to the phase-sensitive detector would then be used to provide only the *sign* of the error.

Summary

In this chapter we have mentioned some of the advances that are being made in primary ground radar systems. As we have seen, the main aims are an improvement in detection probability of signals buried in noise, and improved accuracy of radar measurements by the use of new waveforms and techniques.

Much more could be said about advances on other fronts, notably in aerial design. A lot of effort has gone into the design of large steerable aeriels to improve the operation of long-range search radars. Electronically-scanned aerial arrays (phased arrays) continue to improve and are being used increasingly for specific tasks (see p.238). However, the subject of aeriels is large and would require at least a chapter to itself. We shall therefore leave it for the moment.

There is also research going on into the possible use of the h.f. band (up to 30MHz) and scatter propagation techniques (see AP3302 Part 2) to provide over-the-horizon radars. Conventional microwave radars are confined to line-of-sight operation, with a slight extension if the frequency is reduced to the u.h.f. band (see p.228). Because of this limitation, little time is available to provide warning of the arrival of a missile. If over-the-horizon radars could be developed, warning times could be improved considerably. Little is known about the progress being made in this field but it is clear that at such low frequencies very large aerial systems will be required (possibly 150 ft high) and very high power outputs (probably several hundreds of megawatts peak power).

In closing this chapter it may be worth mentioning a new development of general interest to ground radar technicians—the *automatic adaptive radar*. We saw early in this book (see p.31) that many factors affect the operation of a ground radar system. A ground radar is usually designed with a particular task in mind, and the values of p.r.f., pulse duration, frequency, beam shape, peak power, and so on are usually chosen to provide the best possible results. Even so, many of the factors conflict and the design, at best, is a compromise.

In the adaptive radar, provision is made for changing rapidly many of the transmitter and receiver characteristics to meet varying needs. By changing the radar parameters as required a single ground radar system can produce satisfactory results in many roles. It could be used in one role to provide good probability of detection at long ranges, converting to the role of accurate ground tracking radar as the target closed. With such a system, e.c.m. interference can be avoided by rapidly changing the frequency. The equipment may be used in an m.t.i. mode or as a non-m.t.i. equipment. Frequency agility may be switched in as required. Pulse compression networks may be made available. The possibilities are obviously immense.

However, the resulting equipment would be very complex and would require automatic computer control of all the variables in the system. The computer would be 'told' what was expected of the radar and would adjust the variables to satisfy the requirements. By comparing the receiver output with an 'ideal' output, the computer may be programmed to adjust the variables to provide an output as close to the ideal as possible.

The automatic adaptive radar is in an early stage of development but it is already clear that the advantages it offers are so great that continued development is virtually assured.

SECONDARY SURVEILLANCE RADAR

Introduction

Basic IFF (Identification, Friend or Foe) was introduced during the 1939-45 war to enable cooperating 'friendly' aircraft to be distinguished from uncooperating 'foe'. IFF is a *secondary* radar system in which a ground station transmits interrogating signals to an aircraft which, if suitably equipped, sends appropriate *replies* to the ground station. In effect we have a sentry on guard duty challenging everyone who approaches; only those who are able to give the correct password in reply to the challenge are allowed to pass unmolested. In IFF, the information received at the ground station is normally displayed as an easily recognizable blip on the radar screen alongside the primary radar response, and this blip distinguishes friend from foe.

Basic IFF

The essential requirements of IFF are:

- A transmitter-receiver (interrogator-responder) on the ground.
- A ground station aerial system that produces a rotating beam, narrow in azimuth and wide in elevation.
- A transmitter-receiver (transponder) carried in the aircraft.
- An omni-directional aerial mounted on the aircraft.
- Means of extracting and displaying the information received at the ground station.

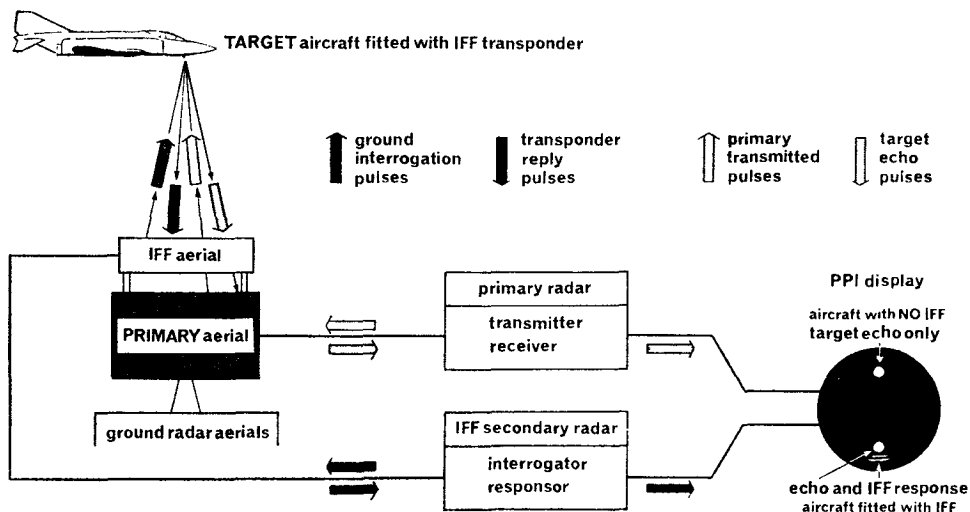


FIG. 1 BASIC IFF/PRIMARY RADAR ARRANGEMENT

Fig. 1 shows that IFF and the primary search radar work in conjunction with each other. In fact, on some ground radar stations, the IFF aerial is mounted on top of the primary search aerial and rotates with it. The operation of the primary radar is as described earlier in this book, the received target echoes being processed to provide the usual blip on the p.p.i. display. The ground IFF equipment radiates a signal which is received by the transponder receiver in the aircraft. The received signal generates a pulse, or a group of pulses, which is used to modulate the transponder transmitter. The reply signals radiated by the transponder are received by the ground IFF responder and processed to produce suitable responses for display side by side with the primary radar echo. The range and bearing of the target are as indicated by the position of the target echo on the p.p.i., and the secondary radar response identifies the target as 'friendly'.

IFF operates within a band of frequencies around 1GHz. The ground interrogator operates at one frequency within this band and the reply signals from the airborne transponder are at a *different* frequency; the frequency separation is of the order of 60MHz. The ground receiver is tuned to the transponder transmitter frequency so that only signals at this frequency are received. This ensures that the reply is due to a transmission, and not an echo, from the target. Secondary radar has other advantages over primary radar; since we are not relying on an echo, the usual limitations on primary radar range do not apply; the signal-to-noise ratio is very much higher in secondary radar and problems caused by clutter are eliminated. Complete line-of-sight coverage may be obtained with low transmitter powers, a typical range being 200nm for interrogator and transponder peak powers of 2kW and 750W respectively.

IFF Modes

From this early concept, merely as a means of distinguishing between friend and foe, IFF has progressed step by step to the stage where it now plays an important part in *air traffic control* (ATC) systems. When used as an aid to ATC, IFF systems are usually referred to as *secondary surveillance radars* (SSR). Let us now consider the various stages in the development of SSR from basic IFF.

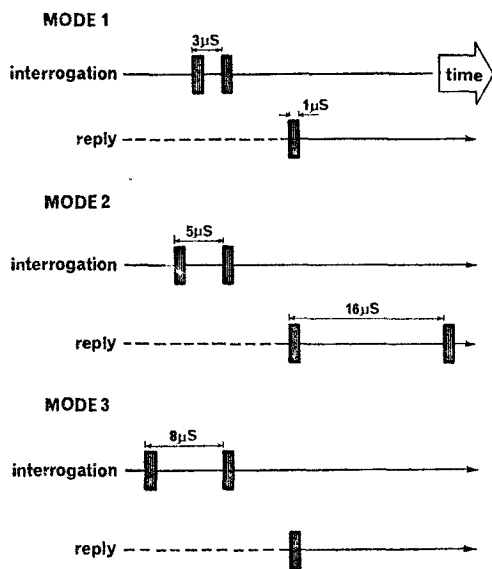


FIG. 2. IFF MODES

The first extension of the basic IFF principles was the introduction of different '*modes*' of operation. In the RAF, the interrogator is designed to operate on any one of three modes of interrogation to meet different operational requirements. In effect, these modes are equivalent to three interrogation codes. The transponder in the aircraft decodes the interrogations received and replies *only to those interrogations* which are in the mode selected on the transponder control unit. The three modes, shown in Fig 2, are as follows:

- Mode 1.* This mode is used for *general identification*, friend or foe. The interrogating signal consists of a pair of $1\mu\text{s}$ pulses with $3\mu\text{s}$ spacing between leading edges. The transponder reply is a single $1\mu\text{s}$ pulse.
- Mode 2.* This is intended for *detailed identification* of specific aircraft. The interrogating signal consists of a pair of

$1\mu\text{s}$ pulses with $5\mu\text{s}$ spacing between leading edges. The transponder reply is a pair of $1\mu\text{s}$ pulses with $16\mu\text{s}$ spacing between leading edges.

- Mode 3.* This is used to specify the *functional class* of an aircraft and has a direct ATC application. In fact, it is the same as the mode used in civil SSR for ATC purposes. The interrogating signal is a pair of $1\mu\text{s}$ pulses with $8\mu\text{s}$ spacing between leading edges. The transponder reply is a single $1\mu\text{s}$ pulse, as for mode 1.

In addition to the three operational modes, the aircraft can use IFF to indicate a *distress* condition. By operating a special emergency switch in the aircraft, the transponder can be set to radiate automatically a sequence of four $1\mu\text{s}$ pulses, the spacing between pulses being $16\mu\text{s}$. This will be radiated in reply to *any mode* of interrogation.

Selective Identification Feature (SIF)

Although some ground radar stations and some aircraft are fitted with equipment that satisfies only the three-mode basic IFF requirements, newer installations and front-line aircraft have an additional facility known as SIF. This enables the system to employ, not only different modes of operation, but also discrete *codes* within each mode. The advantage of a coding capability is that it enables each mode to convey information of a given type more precisely.

When SIF is fitted, the characteristics of the basic *interrogations* for each of the three modes are *unchanged* and remain as shown in Fig 2. However, in the airborne transponder, the SIF equipment causes the normal reply pulse to initiate a *pulse train* consisting of up to 14 pulses (Fig 3).

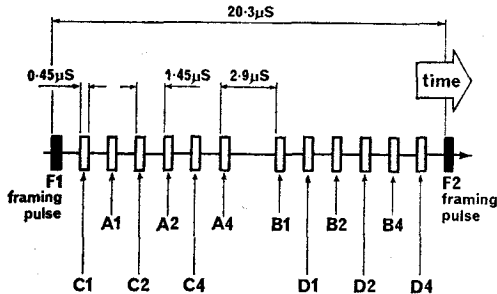


FIG. 3 SIF TRANSPONDER PULSE TRAIN

Any or all of the twelve pulses between the framing pulses may be present in the transponder reply, the actual number depending upon the code selected in a particular mode. Each of the twelve information pulses is given an identifying letter (group A, B, C, or D) and a numerical value (1, 2, or 4).

Note that the groups *do not* run consecutively. Group A and group C pulses are in the first set and are interlaced as shown. Groups B and D are in the second set and are similarly interlaced.

The code consists of four digits, one from each group. Group A pulses provide the *first* digit and, depending upon the number of pulses present in that group, any value between 0 and 7 may be produced (Fig. 4). Group B pulses provide the second digit of the code, again of any value between 0 and 7; group C pulses give the third digit, and group D the fourth. Thus we can have any code combination between 0000 (when only the framing pulses are transmitted) to 7777 (when all pulses are transmitted). An example of a coded reply (1605) showing the pulses needed to make this up is shown in Fig. 4.

If all twelve information pulses are used it is possible to obtain a maximum of 2^{12} or 4096 different codes in *each mode*. At the moment, this is in excess of requirements, and a much smaller number of the available codes is used in practice.

The present usage of the codes is as follows:

Mode 1. Since this is the general identification mode, it will normally employ a *common code* to be used by all friendly aircraft, the code being changed only rarely in peace time. The

The first and last pulses in this train are known as 'framing' or 'bracket' pulses and are spaced $20.3\mu\text{s}$ apart; these two pulses are transmitted in reply to *any mode* of interrogation that the transponder is set up to receive. Twelve pulse positions, each with a designated order of significance, are spaced between the framing pulses, and space is left in the middle for a thirteenth pulse. In effect, we have *two sets* of seven pulses, the spacing between the pulses within each set being $1.45\mu\text{s}$, and between the sets, $2.9\mu\text{s}$.

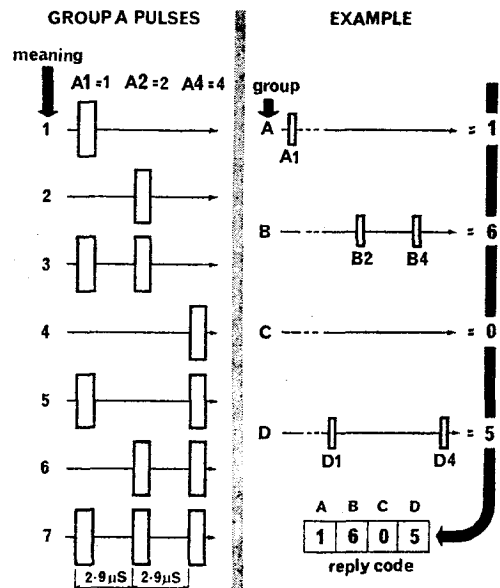


FIG. 4 MEANING OF SIF PULSES

only SIF information pulses used in this mode are A1, A2, A4, B1, and B2, so that a total of 32 (2^5) codes is available (Fig. 5a).

Mode 2. This is the personal identification mode in which aircraft are given an individual code that may be pre-set before flight. At the moment, a *fixed number* of six information pulses is used for each code in this mode. The six pulses are made up of any three pulses from the A and C groups plus any three from B and D. Permutation shows that this provides up to 400 codes.

Mode 3. This mode is used for traffic identification and indicates the functional classification of an aircraft for use by ATC. The code is normally pre-set before flight. The only six information pulses used in this mode at present are any, or all, of the A and B groups. This gives a total of 64 (2^6) codes.

Secondary Surveillance Radar (SSR)

For many years IFF was used solely by military aircraft to provide identity to interrogating radar stations and to enable ground controllers effectively to control the aircraft.

However, the advantages of IFF in air traffic control has become so obvious that civil equipment has now been produced. This equipment is known as secondary surveillance radar (SSR). Because it was produced after much technical experience of IFF it has more refinements.

SSR is similar to IFF in that it can provide general, individual, and functional identifications of an aircraft, superimposed on the positional information obtained from the primary radar. But SSR can also be programmed to provide automatic altitude reporting and other information of great importance to a controller in the ATC system.

Civil SSR is designed to operate on *four modes*, A, B, C, and D, each having a coding capability. Of the three military modes (1, 2, and 3) and the four civil modes, only one of each is compatible: military mode 3 and civil mode A, often referred to as mode 3/A. This is an obvious handicap because it means that RAF radar stations equipped with IFF are able to interrogate civil aircraft only in mode A; and civil ground stations are able to interrogate RAF aircraft only in mode 3. The ultimate aim is a completely common ATC system for both civil and military aircraft in this country. The groundwork of this common integrated system has already been completed and it is clear that, to make it work, all aircraft and ground radar stations will eventually be equipped with SSR. Some RAF aircraft are already so equipped.

The main characteristics of SSR are as follows:

- Ground interrogator transmission frequency: 1030MHz.
- Airborne transponder transmission frequency: 1090MHz.
- PRF: maximum of 450 interrogations per second.
- Interrogations from ground consist of pulse pairs, as in IFF, with spacing between the pulses depending upon the mode in use:

- (1) Mode A: $8\mu\text{s}$ spacing (as for IFF mode 3).
- (2) Mode B: $17\mu\text{s}$.
- (3) Mode C: $21\mu\text{s}$.
- (4) Mode D: $25\mu\text{s}$.

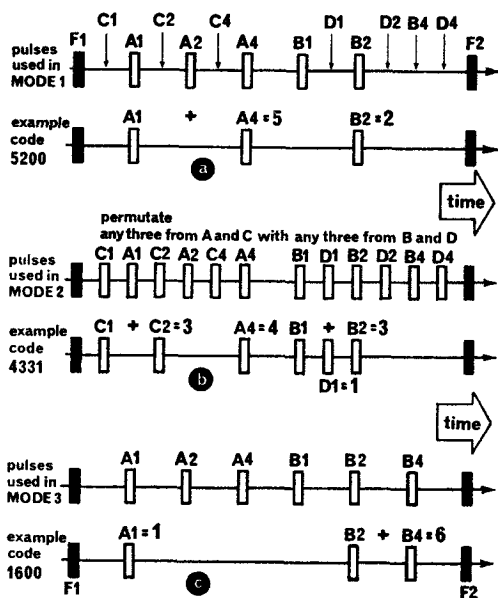


FIG. 5 PRESENT USE OF SIF CODES

- e. Transponder replies on each mode are the same as those described for SIF in Fig. 3, with the addition of a *special identification pulse* inserted, on request, $4.35\mu\text{s}$ after the second framing pulse F2. Thus we have available 4096 codes for each of the four modes of operation.
- f. Ground interrogator pulse duration: $0.8\mu\text{s}$.
- g. Airborne transponder pulse duration: $0.45\mu\text{s}$.
- h. Range: about 200 nautical miles.
- j. Ground interrogator beamwidth: 2.5° in azimuth and 45° in elevation.

SSR Modes

The uses of the various modes in SSR are subject to international agreement. So far it has been agreed that modes A and B shall be used for identity functions and mode C for altitude reporting; the role of mode D has not yet been decided.

Mode A. The present application of this mode in ATC is to provide functional identification, as distinct from individual identification. It will be remembered that military mode 3 has the same application but, whereas only 64 of the available codes are at present used in mode 3, it is intended to make the full 4096 codes available in mode A. The functions expressed in mode 3/A are the identity of the ATC agency handling the aircraft and the identity of the sector or airspace in which the aircraft should be operating. It is possible that other items of information may be added at a later date.

Mode B. This has the same application as mode A and is intended to provide overspill if mode A becomes overcrowded because of its dual military/civil role. In general, aircraft will use *either* mode A or mode B and will not be set up on both.

Mode C. This mode is intended to provide automatic altitude information upon interrogation. One of the main limitations of primary radar is its relative inaccuracy as a height-finder. In the SSR system, the barometric altimeter reading is digitized, the resulting binary digits being available to trigger an appropriate code within mode C. The intention is to provide altitude information in increments of 100 ft, neglecting altimeter inaccuracies; this is essential for accurate control in an ATC system.

Mode D. Although the final role of mode D has not yet been agreed, various suggestions have been put forward. One is that mode D could be used to provide information about the airframe: factors such as wheels up or down, and flaps up or down, could be digitized and used to trigger an appropriate code within mode D. Another suggestion is that mode D could be used to provide *individual* identification of an aircraft for tracking and correlation with the flight plan. In this application an individual four-figure code would be assigned to an aircraft and pre-set before flight, much as it is in military mode 2.

Mode Interlacing

To obtain functional identity of an aircraft from modes A or B, altitude information from mode C, and, say, individual identity from mode D, the interrogator is required to transmit on each mode separately on a *time-sharing* basis. The technique of operating on different modes in turn is known as *mode interlacing*. The ideal mode interlacing programme would be A, B, C, D, A, B, C, D, A — — —. In practice this cannot be achieved because a point is reached where there are insufficient 'hits' by the ground interrogator on the airborne transponder in any mode to ensure satisfactory replies. A *three-mode* interlace is therefore used.

If it is required to interrogate on four modes, a programme such as A, B, C, A, B, C, A, — — — on one aerial rotation, followed by A, C, D, A, C, D, A — — — on the next rotation of the aerial could be applied and repeated. Each mode is selected at *p.r.f. rate*, with each mode *sequence* at aerial rotation rate. The most important modes are included in both sequences (A and C in the example), the information derived from the other modes being received at *half* the data rate.

Sidelobe Suppression (SLS)

The radiation pattern from all aerials contains sidelobes. In primary radar, range is proportional to the *fourth root* of the radiated power and, since the power in the sidelobes is much less than that in the main lobe, the sidelobes cause confusion only at *very short* ranges. This is not the case for secondary radar where the returns to the ground from the target are *not echoes* but transmissions from a transponder; thus, for range, the fourth root relationship no longer applies. Hence sidelobes will be effective at *much longer* ranges in secondary radar. The result is that a transponder, sensitive enough to be triggered at very long ranges by the interrogator main lobe, would also be triggered at medium ranges by the sidelobes.

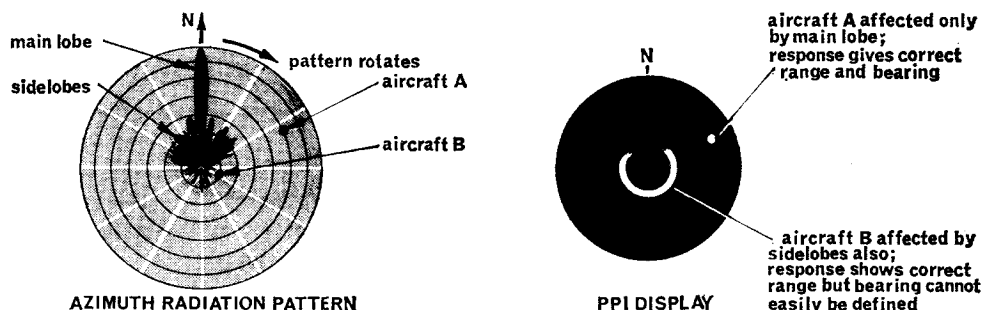


FIG. 6 SIDELobe EFFECT IN SECONDARY RADAR

For an aircraft flying within, and triggered by, a sidelobe, the *correct range* would be displayed on the ground radar p.p.i., but the *bearing* information would be obscured (Fig. 6). As the range decreases the situation becomes progressively worse until, close to the station, the responses on the ground radar p.p.i. may form a *complete ring* and all bearing information is lost.

To eliminate the *effect* of the sidelobes in secondary radar, a system known as sidelobe suppression (SLS) may be incorporated in the ground and airborne equipments. SLS prevents a transponder replying to interrogations except when the aircraft is in the *main lobe* of the interrogator aerial.

The method used is to supplement the interrogator pulse or pulses radiated from the rotating directional aerial by a 'control' pulse which is emitted from an *omni-directional* aerial. The amplitude of the control pulse is such that its signal strength is greater than that of the strongest sidelobe but not as great as that of the interrogator main lobe. Thus, except within a few degrees of the axis of the main lobe, the level of the control signal is *above* that of the interrogator signal (Fig. 7). The airborne transponders are then designed to include a circuit that compares the ratio of the amplitudes of the interrogator and control pulses to decide whether a reply should or should not be sent. Thus, when an airborne transponder is swept by a sidelobe, the sidelobe is swamped by the control pattern and the transponder makes no reply. When swept by the main lobe, the transponder is triggered and replies to the interrogation.

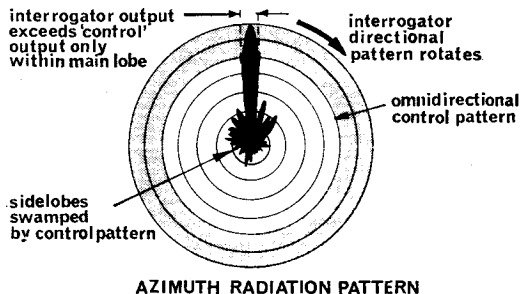


FIG. 7 SIDELobe SUPPRESSION, PRINCIPLE

Two-pulse SLS

There are two forms of sidelobe suppression: the two-pulse, and the three-pulse systems. In the two-pulse system the normal pulse pairs are transmitted by the interrogator, the separation

between the pulses depending upon the mode. The two pulses are referred to as P1 and P3 (Fig 8).

P1 is the *control* pulse which provides the omni-directional radiation pattern. P3 is the *interrogator* pulse which provides the directional radiation pattern. The airborne transponder will reply to properly spaced interrogations of the selected mode when the received amplitude of P3 is above the level shown in Fig. 8; this condition applies when the aircraft is within the narrow main beam of the interrogator aerial. On the other hand, the transponder reply will be suppressed when the received amplitude of P1 exceeds that of P3 by 10db or more; this occurs when the transponder is triggered by a sidelobe.

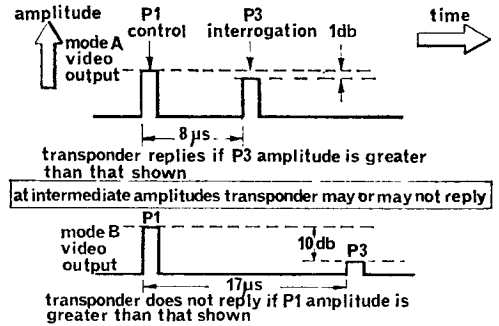


FIG. 8 TWO-PULSE SLS

Three-pulse SLS

In this system, the P1 and P3 pulses, separated by a time interval depending upon the mode, are *both* applied to the *interrogator* aerial to provide the directional radiation pattern. A third pulse, known as the *P2 control pulse*, is generated 2 μ s after the P1 pulse and is used to provide the omni-directional radiation pattern (Fig 9).

The airborne transponder will reply on the selected mode when the received amplitude of P1 is 9db or more above that of the control signal P2; this occurs when the aircraft is within the main lobe. The transponder reply will be suppressed when the received amplitude of the control pulse P2 is equal to or greater than that of P1; this occurs during a sidelobe interrogation.

By international agreement, three-pulse SLS

must be used on mode 3/A. On modes B, C, and D either two-pulse or three-pulse SLS may be used. Most SSR equipments make provision for using both systems.

Defruiting

All secondary surveillance radar stations use 1030MHz for interrogation and 1090MHz for transponder replies. Thus any interrogator can trigger any transponder within range and receive replies in the appropriate mode. This is, of course, a normal requirement of SSR. However, an interrogator can also receive replies from transponders that have been triggered by *other interrogators* and, unless the interrogator can distinguish between replies caused by its own interrogations and those due to others, interference will result. This form of interference is given the name 'fruit' and the process of removing it is called 'defruiting'. All modern ground interrogator-responders include a defruiter.

To reduce interference between ground radar stations, SSR interrogators operate at different p.r.f.s. The defruiter relies on this fact. The defruiter is similar in principle to the delay-line canceller considered in p.502 in that the video output is divided into an undelayed channel and a channel that delays the video by one pulse repetition period (Fig. 10 overleaf). However, in the defruiter, the delay line is normally replaced by a digital store which breaks up the radar range into equal, small range (time) increments.

The information pulses returned from the transponder in reply to a given mode are stored in the digital store at their corresponding time (range) intervals. When replies are received

subsequently during succeeding pulse repetition periods, each pulse in the reply is compared with its stored counterpart and when coincidence in terms of range increment occurs, the signal is allowed to pass to the output. Coincidence can occur only for replies caused by the 'home' station's interrogations at a given p.r.f. Any replies, at different p.r.f.s, received in answer to other interrogations will not be correlated with the store contents and are therefore rejected.

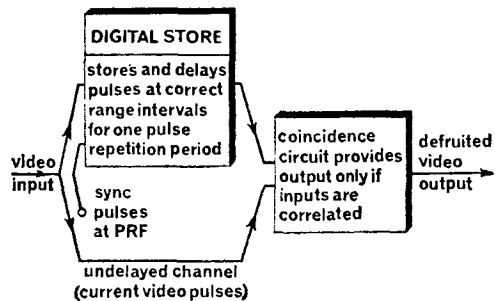


FIG. 10 BASIC DEFRUITER

Ground Interrogator

The preceding paragraphs have discussed, in general terms, the requirements of SSR and the problems to be overcome to achieve the required results. Since the SSR signal route is from the ground interrogator to the airborne transponder, back to the ground responder, it is logical to consider the equipment in that order. A block diagram outlining the stages in a ground interrogator using three-pulse SLS and mode interlacing is shown in Fig. 11 opposite.

Since the secondary radar works in conjunction with the primary radar, and we require the corresponding responses to appear side by side on the p.p.i. display, the input synchronizing pulses to the ground interrogator must be related to the primary p.r.f. By international agreement the maximum permissible p.r.f. in SSR is 450 p.p.s. Up to this maximum the sync pulses will be at the same p.r.f. as the primary radar. If the primary p.r.f. is greater than 450 p.p.s. there are two ways in which the required sync p.r.f. may be obtained: a pulse divider may be used to give a sync input at half the primary p.r.f.; or a suitably gated digital system may be used in which some of the primary radar pulses are inhibited to provide a sync input of 450 p.p.s.

The sync input pulse is amplified and applied to a tapped delay line, the output of which is a train of pulses of the required pulse duration and pulse spacing to produce the various modes. This is illustrated in Fig. 11. The pulse train is applied to the mode selection and mode interlacing circuit. This consists of a series of gates controlled by two separate gating inputs. One input applies gating pulses at the equipment p.r.f. to provide *mode selection* at the *p.r.f. rate*. Thus we have, say, mode A selected for one pulse repetition period, followed by mode B for the next period, and so on. The other gating input consists of a pulse produced for each rotation of the aerial and this operates the selection gates to change the mode *sequence* from, say, A, B, C, A, B, C, A --- on one rotation to, say, A, C, D, A, C, D, A --- on the next.

The P1, P2, P3 pulses in the selected mode are amplified and applied as the modulating signal to the power amplifier, the other input to which is the 1030MHz carrier from a solid-state chain consisting of crystal oscillator, frequency multipliers, and amplifiers. The *power* amplifiers are usually disc-seal triodes.

The modulated r.f. signal is applied to a high-speed switch which is controlled by the P1, P2, P3 pulses in such a way that the P1 and P3 interrogator pulses are applied via a ferrite circulator to the directional aerial, whilst the P2 control pulse is applied via an additional amplifier and circulator to the omni-directional aerial to provide the control pattern. This provides three-pulse SLS. The reason for the additional amplifier in the control chain is that the r.f. output during the control pulse must be greater than that during the interrogator pulses because of the increased power required by an omni-directional aerial relative to a directional aerial. Typical outputs are 2kW on interrogate and 5kW on control.

Airborne Transponder

The basic outline of a typical SSR transponder with three-pulse SLS is illustrated in Fig. 12. The various chains in this equipment are:

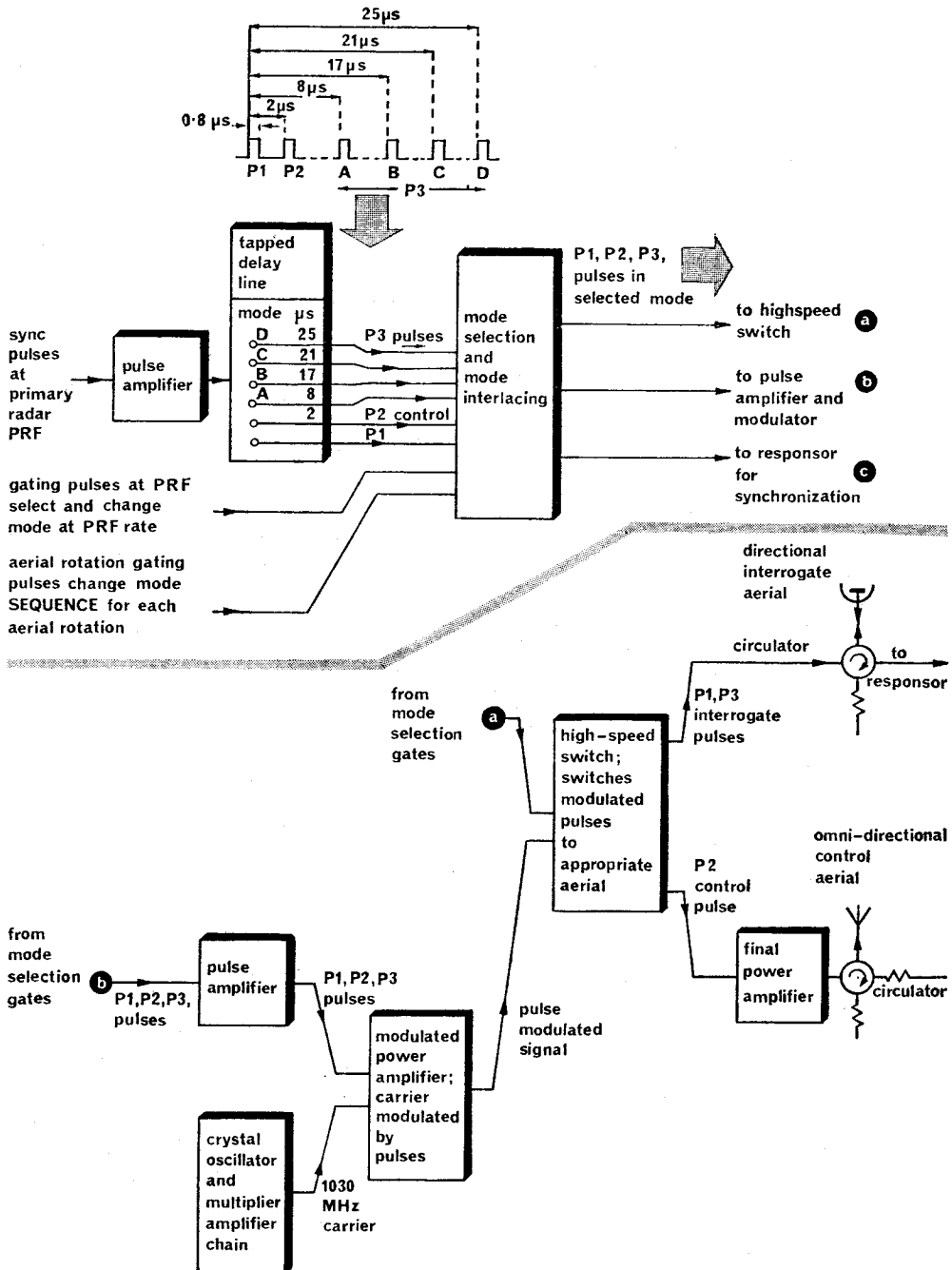
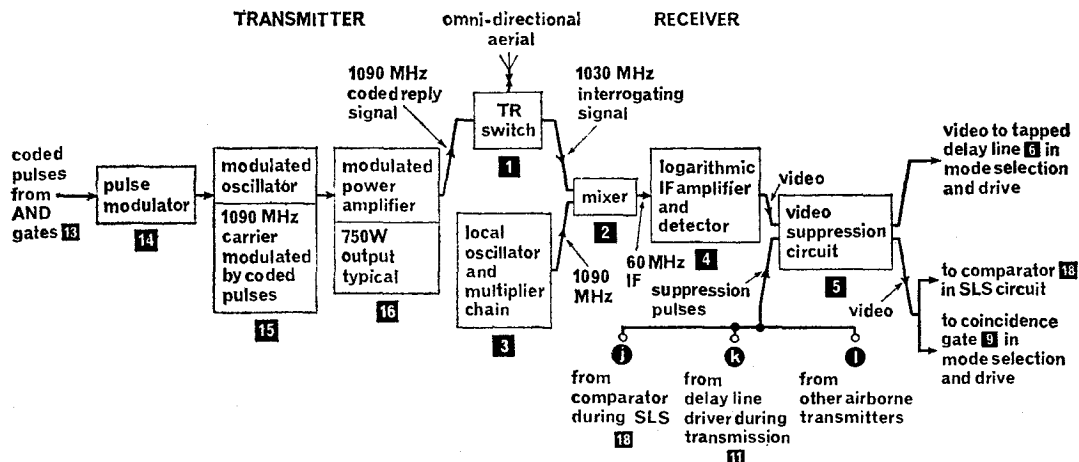
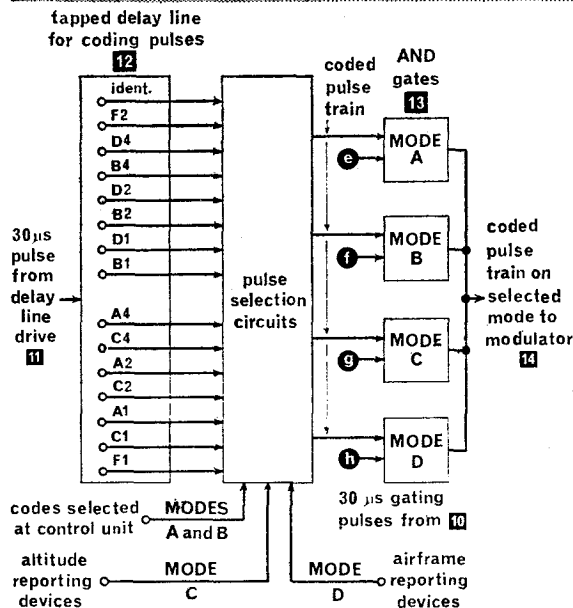
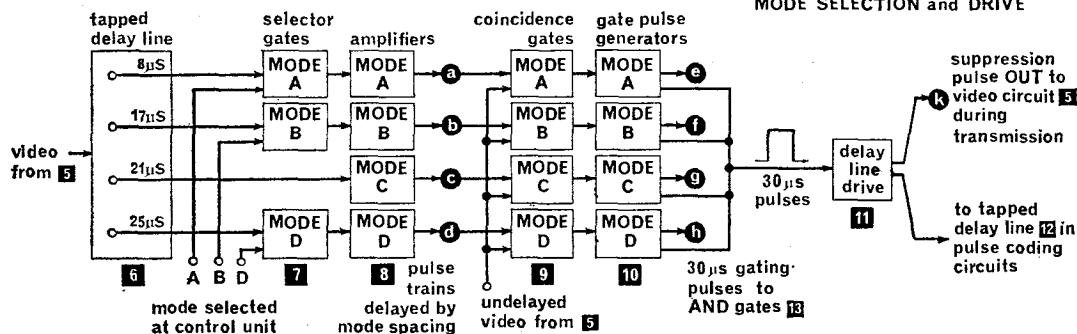


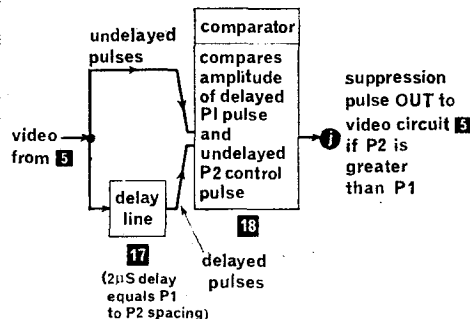
FIG. 11 OUTLINE OF TYPICAL GROUND INTERROGATOR



MODE SELECTION and DRIVE



PULSE CODING CIRCUITS



SIDELOBE SUPPRESSION CIRCUITS

FIG. 12 OUTLINE OF TYPICAL AIRBORNE TRANSPONDER

- a. Transmitter.
- b. Receiver.
- c. Mode selection and drive.
- d. Pulse coding circuits.
- e. Sidelobe suppression circuits.

The incoming interrogating signal is processed in the receiver; the video output is checked for mode of operation and either passed or inhibited; if it is passed, coded reply pulses are produced; the coded pulses are then used to modulate the transmitter to provide the transponder reply. Meantime, the SLS circuits check if the incoming signal is from the main lobe or a sidelobe; if it is a sidelobe interrogation, mode selection is inhibited and no reply is sent.

It will be remembered that the incoming signal from the interrogator may be switching from one mode to another at p.r.f. rate and the sequence of interlacing may also be changing. For ease of explanation we shall assume that the transponder is set up to receive a mode A interrogating signal, although the operation is basically the same for modes B and D also. Mode C is somewhat different; this is the mode used for automatic altitude reporting and the circuit arrangements must be such that a reply will be sent to a mode C interrogating signal *irrespective of the mode selected* at the transponder control unit. Thus if mode A is selected, the transponder will reply to *both* mode A and mode C interrogating signals, but not to inputs in modes B and D.

The receiver is conventional. The incoming interrogating signal at 1030MHz is combined with the output at 1090MHz from the local oscillator to provide the 60MHz i.f. A logarithmic i.f. amplifier with successive detection is used because of the large dynamic range of the incoming signal; the input signal has a large amplitude when the aircraft is close to the ground interrogator, but a much smaller amplitude at greater ranges; the log amplifier attempts to provide a constant amplitude output under these conditions. The video output from the detector is applied to the mode selection circuits through a video suppression circuit which may have suppression pulses *j*, *k*, or *l* applied to it to inhibit further processing.

The video signal, if not inhibited, is applied to a tapped delay line to provide a series of P1, P2, P3 pulses, each pulse train being *delayed in time* by the spacing appropriate to each mode. The pulse trains that are delayed by the mode A, B and D spacings are applied to appropriate selector gates which open only if the mode has been selected at the control unit. The pulse train delayed by the mode C spacing is *not gated* because this channel must be kept available irrespective of the mode selected at the control unit. Thus, if mode A has been selected, an amplified pulse train, delayed in time relative to the current video by the mode spacing ($8\mu\text{s}$), is available at *a*. If the incoming signal is a mode A interrogating signal then the pulse train at *a* has the correct mode A spacing, i.e. $2\mu\text{s}$ between P1 and P2, and $8\mu\text{s}$ between P1 and P3. There is no output at *b* or *d*, but there will be from *c*; for a mode A input signal the output from *c* is inhibited at a later point in the chain.

The mode A delayed pulse train from *a* is applied to a coincidence gate, the other input to which is an *undelayed* pulse train from the video stages. If mode A has been selected at the control unit, and the incoming signal is also on mode A, both delayed and undelayed signals are present at the coincidence gate. Because of the $8\mu\text{s}$ delay in the pulse train from *a*, the *undelayed* P3 pulse coincides in time with the *delayed* P1 pulse (Fig. 13 overleaf). Thus mode A coincidence gate opens to trigger mode A gate pulse generator. The mode A pulse train from *c* is delayed by $21\mu\text{s}$ and does *not* provide coincidence with the undelayed mode A pulses. Thus channel C is inhibited.

It may be seen that if mode A had been selected on the control unit and the incoming signal had been on mode C, no coincidence in time would occur at the input to the mode A coincidence gate because of the difference in mode spacing, and mode A would be inhibited. On the other hand, coincidence would occur at the input to the mode C coincidence gate and mode C channel would be operative.

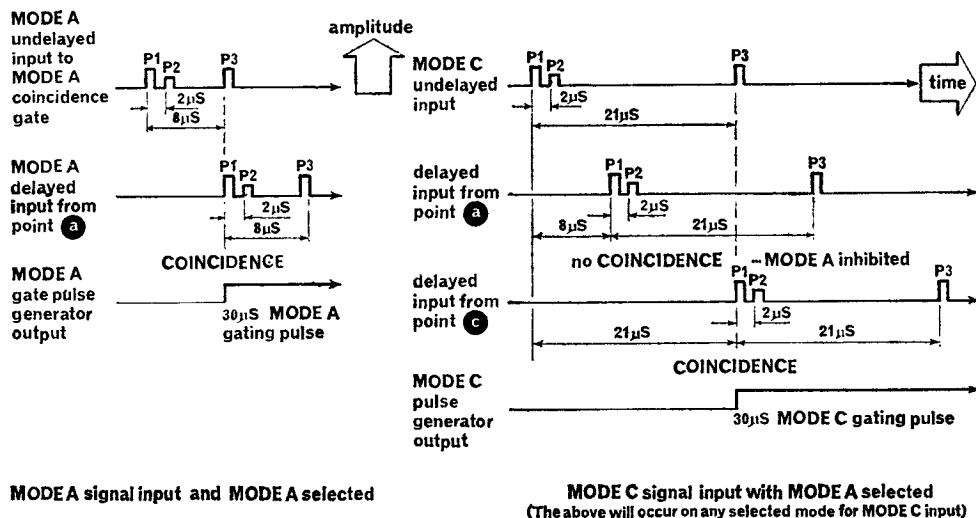


FIG. 13 MODE SELECTION AND REJECTION

Reverting to the mode A operation, when the mode A pulse generator is triggered from the coincidence gate it produces a 30 μ s gating pulse. This pulse is applied to the delay line drive and also as output *e* to the mode A AND gate in the pulse coding circuit. The delay line drive also has two outputs: one output is applied as a suppression pulse *k* to the video suppression circuit to inhibit further operation whilst the current coded reply pulse is being transmitted; the other output is applied to a tapped delay line which generates all the information pulses, F1, C1, A1, etc. In mode A the pulses *selected* from this generated train depend upon the code selected at the control unit. In mode C the pulses selected depend upon the altitude of the aircraft. The selected coded pulse train is applied to the mode A AND gate, which has been opened by the gating pulse from the output *e*, and the output of the AND gate is used to modulate the transmitter to provide the transponder coded reply in mode A.

In three-pulse SLS, the amplitudes of the P1 and P2 pulses are compared to determine whether the interrogating signal initiates from a sidelobe or from the main lobe. At the interrogator the P2 pulse is generated 2 μ s after the P1 pulse, and before an amplitude comparison can be made at the transponder they must be brought into time coincidence. This is achieved by the 2 μ s delay line in the transponder SLS chain. If P2 amplitude is equal to, or greater than, P1 amplitude the interrogating signal is from a sidelobe and the comparator then provides an output *j* which is applied as a suppression pulse to the video suppression circuit to inhibit the P3 pulse and thus prevent coincidence in the mode selection and drive gates.

Ground Responder

A block diagram outlining the required stages in an SSR ground responder is shown in Fig. 14 opposite. The processes in the responder are the opposite of those in the airborne transponder; that is, the received signal is demodulated and the resulting video pulses are extracted in the correct mode and decoded to provide the output to the display. The receiver is conventional and requires no further explanation. Before the video pulses can be processed, it is necessary to ensure that the input signal is in reply to the 'home' interrogations and not to a 'foreigner'. Thus a *defruiter* is necessary. The action has been previously explained and is outlined again in Fig. 14.

The defruited video is applied to the code extractor which effectively checks the validity of the incoming video and *generates* the required information pulses as contained in the transponder reply. Digital techniques may be used for this. In the arrangement shown in Fig. 14, the video pulses are applied to a gate which is opened at the start of each new p.r.f. by gating pulses from

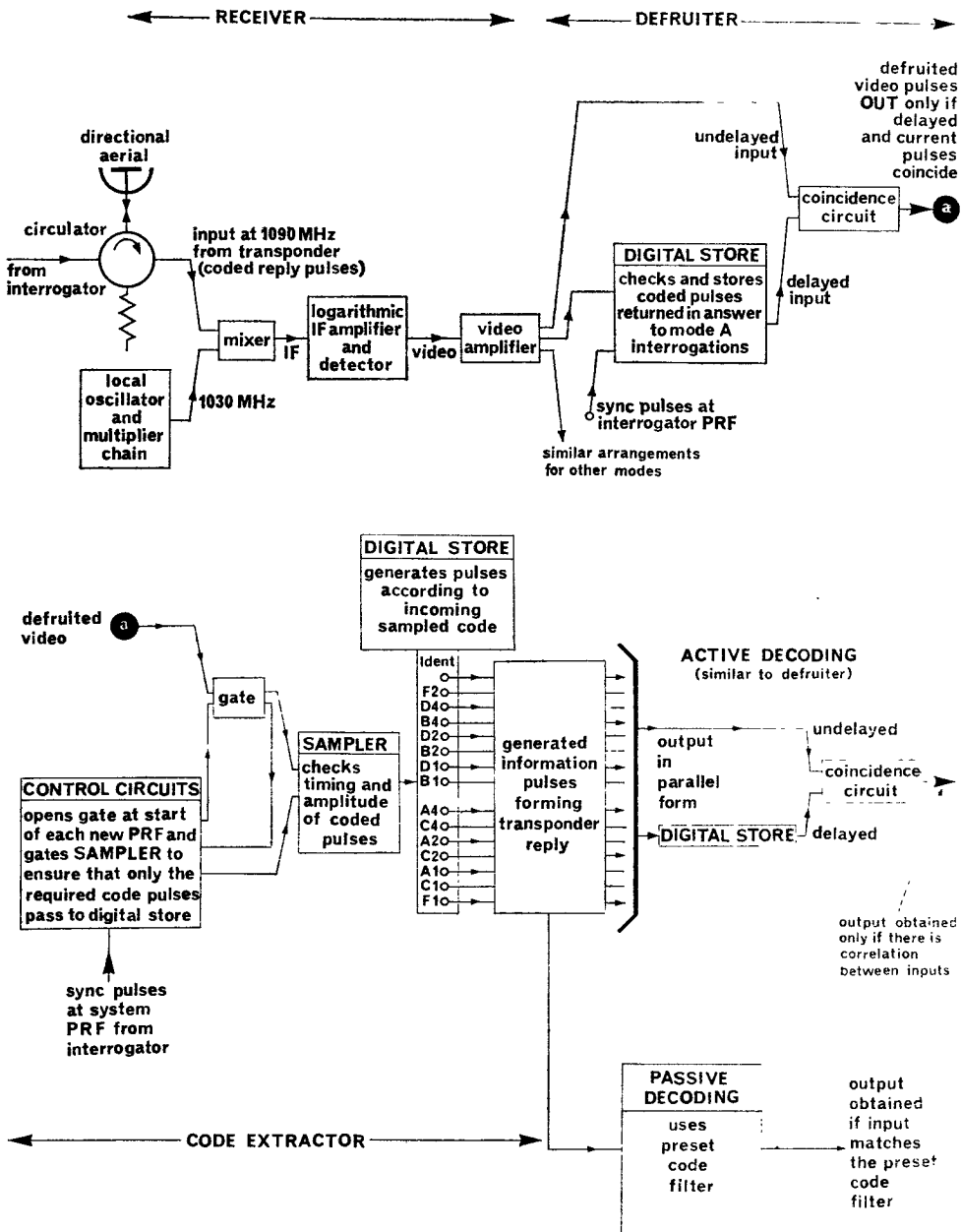


FIG. 14 OUTLINE OF GROUND RESPONSEOR, MODE A ONLY

the control circuits. Gating pulses are also applied to the sampler, the timing being such that only information pulses of the correct duration (and amplitude) are passed to the digital store. The digital store then generates coded pulses in accordance with the sampled input. Each mode has its digital store and the coded pulses generated will be updated in accordance with the sampled input each time the mode is repeated.

After extraction of the codes and generation of the information pulses, it is necessary to check that the information contained in them is worth accepting. There are two ways in which this can be done: *passive* decoding and *active* decoding. With passive decoding the reply to be expected on a given mode of interrogation is *known* so that the controller in the ATC system need only set up an output code filter to accept the expected reply. If the signal generated by the code extractor matches the pre-set code filter an output is obtained; if it does not match, no output is obtained. Thus the passive decoder operates on a simple yes-no basis.

With active decoding nothing is known about the coded reply from the transponder so that an *independent validity check* must be carried out. The method is similar to that in the defruiter: a comparison is made in the coincidence circuit between delayed and undelayed coded pulses. If there is correlation between the two inputs the code is released as output to the display and the validity of the result is assumed to be high. If there is no correlation during one mode repetition period the validity of the output result falls, and if the lack of correlation persists the validity becomes so low as to be unacceptable. Validity is given a 0 to 5 grading, 5 being the highest. In practice active decoding is continued until a validity figure of at least 4 is obtained.

SSR Displays

Different types of display are used in SSR systems depending upon the requirement of the ATC agency. For passive decoding and normal p.p.i. working the SSR signals appear as a raw radar response side by side with the primary echo. The decoded signals from the transponder are under the control of the operator and he can select different responses for each aircraft replying to his interrogations; single-bar, double-bar, and filled-in double-bar are some of the possibilities that may be used to label a particular echo.

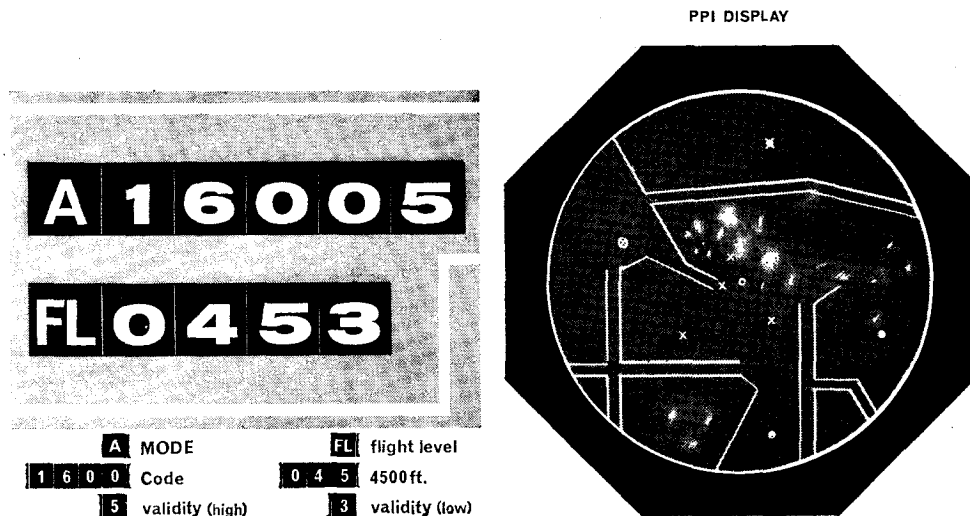


FIG. 15 SSR PPI AND TABULAR DISPLAYS

For active decoding, a tabular display may be used (Fig. 15). With such a display the operator isolates the aircraft in which he is interested by means of a strobe marker on the main p.p.i. display. The output of the decoder is then used to indicate the mode and code of the selected

aircraft on the tabular display. The altitude of the aircraft may also be shown in reply to interrogations on mode C.

The aim for the future is complete automation and integration of all radar-derived information. It is possible to arrange for all target information to be automatically decoded by computer and applied via a character generator to a suitable p.p.i. display. The p.p.i. would indicate the aircraft identity and altitude by alpha-numeric characters in the correct relative position. Fig 16a indicates the responses that may be used in an automatic decoded display and Fig 16b shows an example of a complete display.

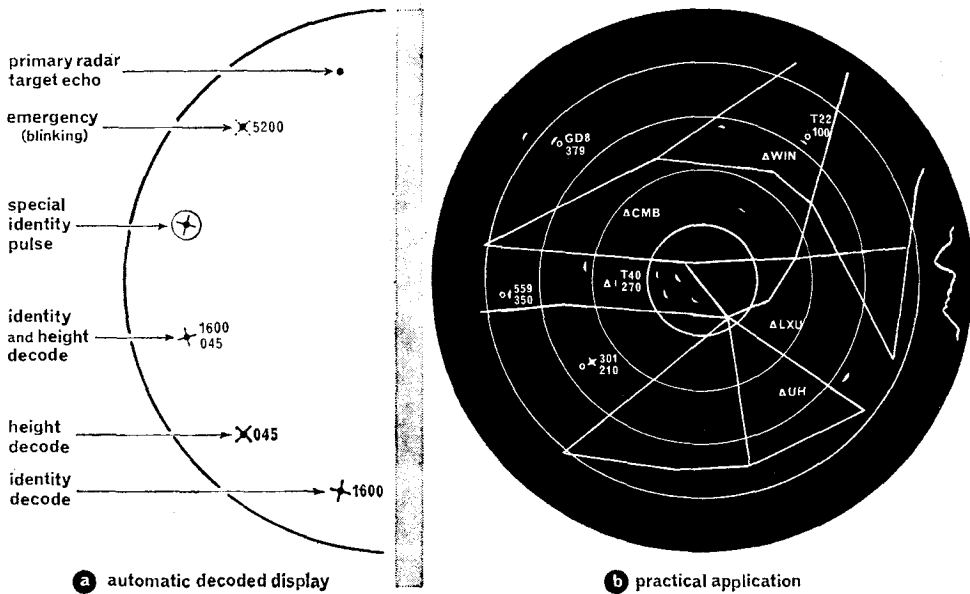


FIG 16. AUTOMATIC DECODED DISPLAYS

It is expected that this type of display will play an important part in the radar defence and control system outlined in Section 8 (see p.538). In such a system, primary and secondary radar data from various remote sites are extracted and transmitted via microwave links to an ATC centre. The data fed to the centre from these sources will include SSR mode, code, range, bearing, and height information. A display processor at the centre collates and consolidates the data from the various sources, converts the range and bearing data into cartesian co-ordinates for correct display of target positions, and converts the SSR data into alpha-numeric characters which are written next to the position plot on the display. Each controller will have a control unit to provide selective display of individual aircraft tracks.

INDEX

This index lists the main contents of this book in alphabetical order; the numbers after the subject give the page references. Most subjects are cross-referred; for example “Varactor diode” under “V” and “Diode, varactor” under “D”. Thus, if a subject cannot be located under one letter, try an appropriate cross-reference.

INDEX

A

- Absorption wavemeter, 253
- Accuracy of radar measurements, 614
- Acoustic delay lines, 504; 592
- Active ECM, 440
- Aerials,
 - beam shape, 33; 34; 337
 - centimetric, 317–339; 481; 490; 526
 - advantages, 317
 - airborne Doppler, 490
 - beam steering, 333
 - beamwidth, 317
 - broadside array, 329
 - cassegrain aerial, 324
 - cosecant-squared reflector, 321
 - dipole feed, 322
 - gain, 318
 - hohorn, 324
 - horn feed, 323
 - janus system, 481; 490
 - lens aerials, 327
 - monopulse multibeam, 337
 - multibeam, 324
 - non-resonant slotted array, 330
 - organ pipe scanner, 326
 - parabolic reflectors, 319
 - pencil-shaped beams, 337
 - polarisation, 318
 - polyrod aerial, 328
 - quarter-wave plates, 326
 - radar data links, 526
 - radiation pattern, 318
 - radomes, 338
 - rear feed with slots, 322
 - resonant slotted array, 330
 - slotted waveguide array, 329
 - squint angle, 331
 - scanning methods, 33
 - scanning speed, 36
 - UHF aerials, 237–247
 - advantages 237
 - electronic scanning, 238
 - frequency scanning, 242
 - methods of feeding, 239
 - multiple-beam arrays, 245
 - non-resonant, 240
 - phase-shifting devices, 241
 - resonant, 240
- AFC (see Automatic frequency control)
- AGC (see Automatic gain control)
- Airborne radar,
 - airborne interception (AI), 48; 595
 - altimeter, 465–476
 - anti-surface vessel (ASV) radar, 595
 - collision-avoidance systems, 607
 - Doppler, 477–498
 - hyperbolic navigation systems, 595
 - identification—friend-or-foe (IFF) 595; 627
 - line scan, 600
 - marine reconnaissance, 601
 - moving target indication (MTI), 600
 - navigation and bombing system (NBS), 595
 - photographic recorder, 597
 - picture quality, SLR, 598
 - Rebecca-Eureka, 595
 - reconnaissance radar, 596
 - sideways-looking radar (SLR), 596–600
 - Tacan, 595
 - terrain-following radar, 602
 - transponder, 634
 - weather radars, 604
- Along-across display, 497
- Altimeters, 53; 465–476
 - aerials, 474
 - barometric, 465
 - block diagram, 488
 - frequency-modulated CW, 54; 466–475
 - modern developments, 475
 - pulse-modulated, 465
 - typical equipment, 474
 - use in automatic landing systems, 474
- Ambiguity, 616
- Amplifiers,
 - amplitron, 234
 - cascode RF, 409
 - crossed-field, 575
 - grounded-grid, 410
 - intermediate frequency (IF), 416
 - klystron, 256; 573
 - klystron/travelling-wave tube, 574
 - neutralised triode, 409
 - parametric, 347–353
 - paraphase, 221–226
 - power, 376
 - pulse, 377
 - radio frequency (RF), 409
 - see-saw, 224
 - travelling-wave tube (TWT), 263; 411; 571
 - video, 419
- Amplitude-comparison monopulse, 625
- Analogue-to-digital conversion, 527–535
 - angle to digits, 528
 - binary coded numbers, 533
 - continuous progression code, 530
 - DC-to-digital coder, 532
 - parallel codes, 530
 - precision switches, 532
 - series codes, 530
 - summation coder, 530
- Angels, 435
- Angle bearing errors, 623
- Angular discrimination, 32
- Anode-coupled multivator, 110–116; 123
- Anode modulation, 374
- Anti-clutter gain control, 430
- Anti-surface vessel (ASV) radar, 595
- Applegate diagram, 256
- Artificial (delay) lines, 365

AP 3302 PART 3

Astable multivibrator, 109–122
ASV (see Anti-surface vessel radar)
Asymmetrical bistable multivibrator, 129
Atmospheric absorption noise, 398
Attenuation absorbers, 442
Attenuators (waveguide), 292
Auto-correlation, 609
Auto-follow strobe circuit, 215
Automatic frequency control (AFC), 308; 414
Automatic gain control (AGC), 417; 428
Avalanche effect, 111
Average power, 28; 36; 39; 362
Azimuth scanning at UHF, 243

B

Backward diode, 554
Backward-wave device, 265
Backward-wave oscillator,
 O type, 265
 M type, 274
 recent developments, 580
Balanced deflection (CRT), 220
Balanced mixer, 306
Balanced TR switching, 312
Bands, radar, 33; 227
Bandwidth, electronic, 260
Bandwidth, receivers, 39; 403
Barometric altimeter, 465
Barrage noise jamming, 440
Beamwidth (aerial), 317
Bearings, radar, 8; 11; 13
Bends, waveguide, 295
Bifilar windings, 379
Bimetallic strip, 387
Binary coded numbers, 533
Binary echo integrator, 614
Bipolar transistors, 545
Bipolar video, Doppler, 503
Bistable multivibrator, 127
 as counter, 174
Blind speeds, 511
Blocking oscillator, 148–152
 as driver stage for modulator, 376
 as frequency divider, 168
Boltzmann's constant, 399
Bootstrap timebase generator, 199
Branch arm TR switching, 311
Bridge rectifier, 385
Broadside array, 329
Buncher cavity, 257
Bunching plane, 256

C

C band, 33
Calibration generators, 146–152
Calibration pips, 143–146
Capacitive tuning of cavity, 252
Carbon anode, 387
Cascode low-noise RF amplifier, 409
Cassegrain aerial, 324

Catcher cavity, klystron, 257
Cathode follower, 92
Cathode-coupled multivibrator, 116–126
Cathode ray tubes,
 electromagnetic, 201
 electrostatic, 175
Cavities, 230; 249 (see also Resonant cavities)
Centimetric aerials, 317–339 (see also Aerials)
Centimetric receivers, 395–443
Characteristic impedance, 367
Charge storage (in diode), 552
Cheese reflector, parabolic, 320
Chip devices, 564
Choke joint, 293
Choke plunger, 294
Circular polarisation, 301; 326
Circular timebase, 205
Circular waveguide, 286 (see also Waveguide)
Circulators, ferrite, 303; 584
Clamping circuits, 95–106
Clipping circuits (see Limiting circuits)
Closed circuit television, 541
Clutter, 427
 angels, 435
 reduction, 455
 weather, 434
Coaxial line resonators, 230
Coaxial mixers, 305
Coho, 502
Coincidence circuits, 158
 as counter, 172
Colpitts oscillator, 412
Collector-coupled multivibrators, 118; 125
Collision-avoidance systems, 607
Communication links, 516
Complementary regenerative pair, 135
Computers for airborne Dopplers, 496
Confusion ECM, 439
Continuous progression code, 530
Continuous wave (CW) radar, 53–60; 445–497
 advantages, 457
 bandwidth, 457
 basic outline, 53
 clutter, 455
 CW Doppler in airborne radar, 483
 Doppler effect, 445
 Doppler radar system, 456
 frequency-modulated CW Doppler, 484
 limitations, 457
 measurement of radial velocity, 454
 minimum range, 458
 noise, 455
 phase-sensitive detector, 450
 power requirements, 457
 range determination, 458
 sign of relative velocity, 452
 simple CW radar system, 447
 superhet CW receiver, 448
 unambiguous Doppler, 452
 velocity tracking gate, 455
Control centre, 515
Converters, parametric, 353
Corners (waveguide), 295

Correlation processes, 609
 Coscant-squared reflector, 321
 Cosmic noise, 398
 Coupling (waveguide), 289
 Count down, 165 (see also Frequency division)
 Counter measures, 434-443 (see also ECM)
 Counting circuits, 169-174
 Cross-correlation, 609
 Crossed-field amplifiers, 573
 CR time constant, 67-75
 Crystal diode, 305; 341
 Cumulative action, 110
 CW radar (see Continuous wave radar)
 Cut-off wavelength, 285
 Cycloid, 269

D **Dartington pair, 341**

Data processing, 518
 DC regulator, 391
 DC restoration (clamping), 95-106
 DC-to-digit coder, 532
 Decca navigation system, 595
 Deception ECM, 439-443
 Decoys, 441
 Defence, radar system, 515
 Definition (of display), 37
 Deflection defocus, 220
 Deflection sensitivity, 180
 Defruiting, 633
 Degenerate parametric amplifier, 348
 Delay circuit (power supply), 388
 Delay lines (electrical), 365-372; 591
 electron beam, 592
 superconducting, 591
 Delay lines (mechanical), 504; 592
 acoustic, 504; 592
 ferrite, 593
 mercury, 505
 quartz, 506
 Delay line canceller, 507
 Delay line modulators, 377-382
 Detection, 3
 Detectors,
 basic circuit, 418
 coherent, 613
 cycle-counting, 612
 envelope, 612
 logarithmic, 432; 612
 Developments in primary radar, 609-626
 Developments in secondary radar, 627-641
 Differential phase-shift circulator, 584
 Differentiating circuit, 72
 Digital data transmission, 517
 Digital phase-shifter, 589
 Diode,
 backward, 554
 crystal, 305; 341
 hot carrier, 556
 IMPATT, 558
 limiters, 83-88

PIN, 343; 558 (see also PIN diode)
 noise generator, 403
 Read, 558
 silicon avalanche, 558
 step-by-step counter, 170
 step recovery, 552
 tunnel, 342; 553 (see also Tunnel diode)
 varactor, 344; 550 (see also Varactor diode)
 Directional coupler, 296
 Disc seal triode, 231; 409
 Discriminator, PRF, 438
 pulse duration, 437
 Dish reflector, 319
 Displays,
 along-across, 497
 closed circuit television, 541
 height-range, 22
 line scan, 600
 plan position indicator (PPI), 21; 205
 radar data, 520; 537
 secondary surveillance radar (SSR), 640
 sector, 22
 synthetic, 537
 tabular, 529
 types, 21
 video map, 539
 Distortion in CRTs, 219
 Dither-tuned magnetron, 578
 Division, frequency, 165
 DOFIC, 566 (see also Integrated circuits)
 Doppler effect, 55-60; 445-447
 airborne radar, 59; 477-498
 aerials, 488
 computers, 496
 CW Doppler, 483
 drift, measurement of, 493
 electronic tracking, 495
 frequency tracking, 490
 FMCW Doppler, 484
 groundspeed, measurement of, 477
 height holes, 488
 Janus system, 481
 operating frequency, 478
 phonic wheel oscillators, 491
 pulsed Doppler, 486
 sea bias, 488
 single aerial, 479
 single-line tracking, 495
 spectrum, 488
 twin-line tracking, 491
 two beams, 479
 Doppler radar system, 456
 ground radar, 58; 445
 principle, 55; 445
 shift, Doppler, 55
 Diode step-by-step counter, 170
 Double-cavity klystron, 256
 Double delay line canceller, 510
 Down-converter (parametric), 353
 Drift, measurement of, 493
 Drift space, 257

AP 3302 PART 3

Dummy load (waveguide), 293
Duty cycle, 28
Duty factor, 28

E

Eccles-Jordan bistable circuit, 127; 174
Echo box, 253
ECM (see Electronic counter-measures)
Electron beam delay lines, 592
Electron-coupled multivibrator, 116
Electronic bandwidth (reflex klystron), 260
Electronic counter-measures (ECM), 439–443
 detection of, 602
Electronic scanning, UHF, 228
Electronic switches, 155
Electronic tracking, Doppler, 495
Elevation scanning at UHF, 243
Emitter-coupled multivibrator, 120; 126
Energy bands, 355
Energy diagram (maser), 355
Eureka, 50
Extraction of radar information, 614

F

Factors affecting radar operation, 29–31
Faraday rotation circulator, 585
Ferrite devices, 302; 581–591
 circulator, 303
 delay lines, 593
 effect in microwave field, 582
 electron spin, 581
 gyrators, 302; 589
 gyromagnetic resonance, 302; 583
 isolators, 303; 587
 linewidth, 582
 phase-shift circulators, 584
 phase-shifters, 589
 resonance, 302; 583
 TR switches, 315
 YIG filters, 590
Field-effect transistor (FET), 548
Fixed coil PPI display, 206
Flange joint (waveguide), 293
Flexible waveguide, 295
Flip-flop (see Monostable multivibrator)
Floating PPA, 224
FMCW Doppler in airborne radar, 484
Forward-wave device, 263
Foster-Seeley discriminator, 416
Free-running (astable) multivibrator, 107–116
Frequency agility, 578; 622
Frequency bands, radar, 33; 227
 choice of, 230
Frequency changing, 305
Frequency division, 165
Frequency of operation, radar, 32; 39
Frequency scanning at UHF, 242
Frequency tracking, Doppler, 490
Fruit, 633
Full-wave rectifier, 385

G

Gain, aerial, 318
Garnet, 591
Gas detection (of submarines), 602
Gas discharge tube as noise generator, 403
Gas-filled rectifiers, 384
Gated AGC, 428
Gating circuits, 157
GEE navigation system, 595
Grass, 21
Grid current clamping, 102
Grid modulation, 375
Ground controlled interception radar (GCI), 43
Grounded grid RF amplifier, 410
Ground interrogator (SSR), 634
Groundspeed measurement, 492
Ground responder (SSR), 638
Group velocity, 285
Guide wavelength, 284
Gunn effect devices, 559–562
Gyrators, ferrite, 302; 589
 digital phase shifters, 589
 TR switching, 315
Gyromagnetic resonance, 302; 583

H

Half-wave rectifier, 385
Hard valve modulator, 160
Hard valve timebase generator, 185
Harmonic content of a square wave, 64
Hartley oscillator, 373
Haystack waveform generator, 203
Height, 11–13
Height holes, 488
High-frequency compensation, 421
Hoghorn, 324
HO1 mode, 283
Horn (waveguide), 323
Hover meter, 498
Hybrid circuits (microcircuits), 564
Hybrid klystron-TWT, 574
Hybrid ring, 299; 408
Hybrid T junction, 299
 balanced mixer, 307
 circulator, 304
 monopulse bridge, 624
 TR systems, 314; 408
Hydrogen thyratron, 379
Hyperbolic navigation, 595

I

Ideal receiver, 401
Idler circuits, 349–353
IFF, 595; 627
Ignitron, 162
 as modulator trigger valve, 379
IMPATT devices, 558
Indicators, 21 (see also Displays)
Induced grid noise, 400

Induction regulator, 389
 Infra-red detection, 602
 Instantaneous AGC, 429
 Integrated circuits (microcircuits), 563
 Integrating CR circuits, 74
 Integration of echo pulses, 613
 Interception equipment, airborne (AI), 595
 Interdigital line, 266
 Interference, 437
 absorbers, 442
 methods for reducing, 439
 PRF discriminator, 438
 pulse duration discriminator, 427
 Interlock circuits, 394
 Interrogator, ground, 24; 634
 Irises, waveguide, 291
 Isolators, ferrite, 303; 587

J

J band, 33
 Jamming, 439-443
 Janus system, 481
 Johnson noise, 398
 Joints, waveguide, 293

K

K band, 33
 Klystron,
 amplifier, 256
 double-cavity, 256
 multicavity, 257; 364; 573
 Applegate diagram, 256; 259
 klystron-TWT amplifier, 574
 recent developments, 573
 reflex, 258; 579
 as local oscillator, 413
 velocity modulation, 255

L

L band, 33
 Lasers, 359
 LC circuits, 78
 Lecher bars, 231
 Lecher bar oscillator, 364
 Lens aerials, 327
 Limiting circuits, 83; 94
 Linear frequency-modulated pulse compression, 621
 Line scan, 600
 Linewidth, ferrite, 582
 Link receiver, 524
 Link transmitter, 523
 Load, dummy (for waveguide), 293
 Local oscillator, 412
 Logarithmic receiver, 432
 Long CR circuit, 73
 Long LR circuit, 78
 Long-tailed pair, 222

Loop coupling, waveguide, 289
 cavity, 251
 Loran navigation system, 595
 Low-frequency compensation, 421
 LR time constant, 77

M

Magic-T junction, 299; 408
 Magnetic anomaly detector (submarines), 602
 Magnetron, 16; 233; 269-277; 577
 application, 364
 dither-tuned, 579
 mechanical tuning, 578
 modes of oscillation, 273
 multicavity, 271
 operating details, 274
 recent developments, 577
 split anode, 270
 strapping, 272
 voltage-tuned, 578
 Man-made noise, 398
 Marine reconnaissance, 601
 Masers, 355-359
 application, 410
 Master timing unit, 16; 33
 Matched-filter receiver, 405; 439; 610
 M-type carcinotron, 274
 Measurements, radar, 614
 Measurements, waveguide, 296
 Medium CR circuit, 72
 Medium LR circuit, 78
 Memory, Doppler tracker, 493
 Mercury arc rectifier, 387
 Mercury delay line, 505
 Methods of feeding UHF aerials, 239
 Microelectronics, 562-567
 Microstrip, 343
 Microwave link, 523 (see also Radar data links)
 Microwave planar triodes, 570
 Microwave semiconductor devices, 543-567
 Miller effect, 191
 in IF amplifiers, 418
 Miller timebase generator, 191
 Miller-transitron, 194
 Mixers, 305; 410
 balanced mixing, 306
 coaxial, 305
 low radar frequencies, 411
 microwave single crystal, 412
 waveguide, 306
 Modes, waveguide, 283
 Modulators,
 bifilar windings, 379
 delay-line modulators, 377
 driver-power amplifier modulator, 375
 grid modulation, 375
 hard valve series modulator, 374
 hold-off diode, 380
 low-power anode modulator, 374
 PIN diode, 344

AP 3302 PART 3

- pulse-driven driver, 371
- pulse-forming network, 373
- pulse power amplifier, 377
- pulse radar transmitter, 382
- pulse transformer, 379
- resonant choke charging, 380
- triggering valves, 379
- Monopulse radar, 337; 442; 624
- Monostable multivibrator, 123–127
- Moving target indicator (MTI), 30; 499–514
 - airborne, 600
 - bipolar video, 503
 - blind speeds, 511
 - coho, 502
 - delay line canceller, 507
 - delay lines, 504
 - double delay line canceller, 510
 - generation of PRF, 508
 - magnetron type, 503
 - MO-PA type, 502
 - principle, 500
 - response of delay line canceller, 509
 - selective MTI, 514
 - staggered PRF, 512
 - stalo, 503
 - type A display, 501
 - unipolar video, 502
- M-type backward-wave oscillator, 274
- Multir strobe circuit, 211
- Multicavity klystron, 257
- Multicavity TR cell, 310
- Multiple beam arrays at UHF, 245
- Multivibrators, 107–133
 - astable, 110–122
 - anode-coupled, 110
 - cathode-coupled, 116
 - collector-coupled, 118
 - control of frequency, 112
 - control of symmetry, 112
 - electron-coupled, 116
 - emitter-coupled, 120
 - free-running, 110
 - synchronisation, 115
 - uses, 121
 - basic facts, 107
 - bistable, 127–131
 - asymmetrical, 129
 - Eccles-Jordan, 127
 - Schmitt trigger, 131
 - transistorised, 129
 - counter, 174
 - frequency divider, 166
 - monostable, 123–127
 - anode-coupled, 123
 - cathode-coupled, 126
 - collector-coupled, 125
 - emitter-coupled, 126

N

- Narrow-band jamming, 440
- Navigation and bombing system, (NBS), 595

- Negative clamping, 98
- Negative feedback (in video stages), 420
- Negative resistance parametric amplifiers, 353
- Neon stabiliser, 390
- Neutralised triode, 409
- Noise, 398–403
 - atmospheric, 398
 - cosmic, 398
 - diode noise generator, 402
 - effect of RF amplifier, 401
 - effect on displays, 20
 - external, 29; 398
 - factor, 31; 400; 417
 - figure, 31; 400; 417
 - gas discharge tube noise generator, 402
 - induced grid noise, 400
 - Johnson, 398
 - man-made, 398
 - measurement of noise factor, 401
 - microwave devices, 355
 - noise power, 399
 - partition noise, 400
 - semiconductor noise, 400
 - signal-to-noise ratio, 398; 400; 609
 - shot, 399
 - thermal, 398
 - transmitter, 455
- Non-degenerate parametric amplifier, 349
- Non-resonant UHF aerial, 240

O

- O band, 33
- Open-circuit delay line, 368
- Optical line scan, 600
- Organ pipe scanner, 326
- Oscillators,
 - backward-wave, M-type, 274
 - O-type, 265
 - blocking, 148
 - Colpitts, 412
 - Hartley, 373
 - klystron, reflex, 258; 573
 - lecher bar, 364
 - magnetron, 16; 233; 269; 577
 - multivibrators, 107–133
 - O-type carcinotron, 265
 - parametric, 353
 - push-pull lecher bar, 364
 - ringing, 80; 146
 - square-wave (see Multivibrators)
 - squegging, 373
 - transmitter, 16; 24
 - ultra audion, 364

Overswing diodes, 381

P

- Parabolic reflectors, 319
- Parallel codes, 530
- Parallel limiters, 85
- Paralysis, receiver, 422
- Parametric amplifier, 347–353
 - application, 410

Parametric converters, 353
 Parametric oscillator, 353
 Paraphase amplifiers (PPA), 221–226
 Partition noise, 400
 P band, 33
 Passive ECM, 440
 Pedestal waveform generator, 203
 Pencil-shaped beam, 337
 Periodic focusing, 264
 Phantastron, 137; 197
 Phase-sensitive detector, 450
 Phase-shifter, ferrite, 302
 digital, 589
 Phase-shifting devices at UHF, 241
 Phase splitter, 221
 Phase velocity, 285
 Phonic wheel oscillator, 491
 Photographic recorder, 597
 PIN diode, 343; 558
 Planar triodes, 231; 364; 570
 Planks constant, 355
 Plan position indicator (PPI), 21; 205
 Platinotron, 275
 Point contact diode, 341
 Polarisation, 318
 Polyrod aerial, 328
 Population inversion, maser, 356
 Positive clamping, 97
 Power,
 mean, 28; 36; 39; 362
 measurement, waveguide, 298
 peak, 27; 31; 362
 noise, 399
 supplies (for transmitter), 371; 383–394
 PPA (see Paraphase amplifier)
 Precession (in ferrites), 302; 582
 Precision switches, 532
 PRF (see Pulse repetition frequency)
 PRF discriminator, 438
 PRF meter, 172
 Primary radar, 2; 23
 Printed circuits, microwave, 287
 Probe coupling, cavity, 252
 waveguide, 289
 Propagation, 526
 Protection circuits (power supply), 394
 Puckle timebase generator, 187
 Pulse compression, 371; 617–622
 comparison with short pulse, 622
 filter, 377
 linear frequency-modulated, 621
 random frequency-modulated, 620
 ratio, 621
 phase-coded, 618
 techniques, 371
 Pulsed Doppler, 486
 Pulse radar transmitter, 361–372; 383
 Pulse duration, 15; 27; 63; 362
 Pulse duration discriminator, 437
 Pulse-forming network, 368; 373
 Pulse generators (see particular type)

Pulse length (see Pulse duration)
 Pulse-modulated radar, 15–51; 361; 372
 Pulse-modulated radar altimeter, 465
 Pulse modulation, 15–51
 Pulse repetition frequency, 15; 27; 35; 63; 362
 staggered, 512
 Pulse shape, 64
 Pulse shaping circuits, 88–94
 Pulse transformer, 373; 379
 Pulse width (see Pulse duration)
 Pump oscillator, 348–353

Q

Q band, 33; 230
 Quarter-wave plates, 325
 Quartz delay line, 506
 Quench circuit, 157

R

Radar,
 accuracy of measurements, 614
 altimeter, 465–476 (see also Altimeter)
 ambiguity of measurements, 616
 applications, 40
 bands, 33; 227
 CW, 445 (see also Continuous wave radar)
 developments, primary radar, 609–626
 secondary radar, 627–641
 factors affecting operation, 29
 factors affecting performance, 29
 fixed-frequency, 622
 frequency-agile, 622
 IFF, 595; 627
 information extraction, 614
 introduction, 1–60
 monopulse tracking, 337; 442; 624
 primary, 2; 23
 pulse-modulated, 15–51; 361; 372
 range, improvement of, 623
 receivers, 234; 395–443 (see also Receivers)
 secondary, 2; 24; 50
 secondary surveillance (SSR), 627–641
 selective identification feature (SIF), 629
 timebase requirements, 188
 transmitter, 361–372 (see also Transmitters)
 weather, 604
 Radar data links, 515; 523–541
 analogue-to-digital conversion, 527–535
 communication links, 516
 data processing, 518
 digital data transmission, 517
 displays, 520; 537
 propagation, 526
 receiver, 524
 repeaters, 524
 secondary surveillance radar, 519
 solid state, 526
 transmitter, 523

AP 3302 PART 3

- Radar defence, 515
 - Radar range, improvement, 623
 - Radial timebase, 205
 - Radial velocity, 454
 - Radiation pattern, 318
 - Radio altimeter, 465–476 (see also Altimeter)
 - Radomes, 338
 - Railing, 437
 - Random FM pulse compression, 620
 - Range gate stealer, 442
 - Range, radar, 3
 - Rat race, 299
 - Read diode, 558
 - Rebecca-Eureka, 50
 - Receivers,
 - automatic frequency control (AFC), 414
 - bandwidth, 39; 403
 - block diagram, 395; 407
 - centimetric, 395–443
 - characteristics, 395
 - clutter, 427
 - detector, 418; 612
 - external noise sources, 398
 - gain, 397
 - IF amplifiers, 416
 - interference, 437–443
 - internal noise sources, 399
 - jamming, 437–443
 - local oscillators, 412
 - logarithmic, 432
 - mixers, 410
 - noise, 397–406
 - noise factor, 400
 - paralysis, 422
 - RF amplifiers, 409
 - sensitivity, 31
 - signal-to-noise ratio, 398; 400; 609
 - solid state receiver, 422
 - UHF, 234
 - video amplifiers, 419
 - Reconnaissance radar (airborne), 596
 - marine, 601
 - sideways looking, 596
 - Rectangular waveguide, 280–286
 - Rectifiers, 384–387
 - Reduction of clutter, 427–435; 623
 - in CW radar, 455
 - Reduction of radar echoing area, 442
 - Reflection, 2
 - Reflex klystron, 258
 - recent developments, 579
 - Regenerative pair, 135
 - Regulation, voltage, 389–393
 - Relaxation oscillator (see particular type)
 - Relaxation time (maser), 356
 - Repeater jamming, 442
 - Repeaters, data transmission, 524
 - Repeller, 258
 - Reset, bistable, 130
 - Resolution, 616
 - Resolver synchro, 207
 - as a phase-shifter, 213
 - Resonance isolator, ferrite, 303
 - Resonant cavities, 230; 249–254
 - absorption wavemeter, 253
 - capacitive tuning, 252
 - coupling, 251
 - development, 250
 - echo box, 253
 - inductive tuning, 252
 - loop, 251
 - probe, 252
 - slot, 252
 - uses of, 252
 - Resonant choke charging, 380
 - Resonant UHF aerial, 240
 - Responsor, 627; 638
 - RF amplifier, 409
 - Rhumbatron, 261; 309
 - Ringing oscillator, 146
 - Roller map, 497
 - Rotating coil PPI timebase, 206
 - Rotating joint, waveguide, 294
 - Rotating reflectors, 441
- ## S
- Sanatron, 139; 197
 - Saturable reactor (as voltage stabiliser), 390
 - S band, 33; 230
 - Scanning, 18 (see also Aerials)
 - Schmitt trigger bistable circuits, 131
 - Scott-connected transformer, 385
 - Sea bias, 488
 - Search radars, 44
 - aerials, 334
 - long range, 44
 - medium range, 46
 - Secondary radar, 2; 24; 50
 - Secondary surveillance radar (SSR), 627–641
 - See-saw amplifier, 224
 - Selective identification feature (SIF), 629
 - Semiconductor devices, 543–567
 - Semiconductor noise, 400
 - Sensitivity time control, 430
 - Series codes, 530
 - Series limiters, 84
 - Series modulator, 374
 - Series stabiliser, 391
 - Set, bistable, 130
 - Short circuit delay line, 369
 - Short CR circuit, 71
 - Short LR circuit, 78
 - Shot noise, 399
 - Shutters, waveguide, 315
 - Side lobe suppression (SLS), 632
 - Sideways looking radar (SLR), 596
 - Signal-to-noise ratio, 398; 400; 609
 - Silicon avalanche diode, 558
 - Silicon controlled rectifier, 384
 - Silicon monoxide, 563
 - Single line Doppler tracking, 491
 - Skin effect, 279, 286
 - Slot coupling, cavity, 252
 - waveguide, 289

Slotted waveguide array, 329–334; 490
 Slow wave structure, 263
 Snow, 21
 Soft rhumbatron TR switch, 309
 Solid state frequency multiplier, 347
 Solid state microwave data link, 526
 Solid state receiver, 422
 Sonobuoys, 602
 Spectrum, radar, 404; 488
 Speed-up capacitors, 129
 Spontaneous emission, 356
 Spot-noise jamming, 440
 Squarer, diode, 87
 Square wave generators (see particular type)
 Square waves, 61–141
 amplitude, 62
 definition, 61
 frequency, 63
 harmonic content, 64
 period, 63
 polarity, 62
 Squegging, 373
 Squint angle, 331; 489
 Stabilisation, voltage, 389–393
 Stabilotron, 277
 Staggered PRF, 512
 Stagger tuning, 417
 Standing wave ratio indicator, 297
 Stalo, MTI, 503
 Step diode, timebase, 203
 Step recovery diode, 552
 Stimulated emission, 356
 Striker anode, 387
 Strobe pulse circuits, 209–218
 Stubs, waveguide, 290
 Summation coder, 531
 Superconducting delay lines, 591
 Surge impedance, 367
 Swept gain, 430
 Switching circuits, 155–163
 Synchro unit, 21
 Sync pulses, 21
 Synthetic displays, 537

T

Tabular displays, 539; 640
 TACAN, 595
 Tapped delay line integrator, 614
 Target dimensions, 30
 Television, closed circuit, 541
 Terrain-following radar, 602
 Thermal delay switch, 387
 Thermal noise, 398
 Thin-film circuits, 563
 Third time round signal, 506
 Three-phase connections, 385
 Threshold of detection, 611
 Thyatron switch, 160
 as modulator, 379
 Thyristors, 391

 as rectifier, 384
 as stabiliser, 392
 Tilting motors, 20
 Timebase principles, 24; 175–189
 electromagnetic CRT, 201–208
 electrostatic CRT, 177–200
 Time constant, 67–81
 T-junctions, waveguide, 298
 Torus reflector, parabolic, 321
 Tracking, Doppler, 490–496
 Tracking radars, 429; 441; 624
 Transformers, power supply, 384
 pulse, 371; 379
 Transistor,
 field effect (FET), 548
 frequency limitations, 544
 frequency multiplication, 547
 metal base, 547
 metal oxide (MOST), 549
 microwave, 545
 unijunction, 152
 Transistor counter, 172
 Transistorised Miller timebase, 198
 Transistron, 136
 Transit time, 231
 Transmit-receive (TR) switching, 18; 308; 407
 Transmitted waveform, 616
 Transmitter blocking (TB) switch, 310
 Transmitters, 15; 361–372
 high-power oscillator, 362
 master oscillator—power amplifier, 363
 mean power, 362
 peak power, 362
 power supplies, 371; 383
 PRF, 362
 pulse compression, 617
 pulse duration, 362
 radar data, 523–541
 range, 361
 transmitter valves, 363; 569
 Transponder, 24
 airborne, 634
 repeater, jamming, 442
 Transverse electric (TE) wave, 283
 Transverse electromagnetic (TEM) wave, 279
 Trapezium distortion, 220
 Travelling wave maser, 357
 Travelling wave tubes, 234; 263–267
 recent developments, 571
 TWT-klystron amplifier, 574
 Trigratron, 162
 as frequency divider, 168
 as modulator trigger, 379
 as timebase generator, 180
 Trigger circuits, 158
 Triggered multivibrators (see Multivibrators)
 Tunnel diode, 342; 553
 amplifier, 343; 410; 554
 frequency divider, 168
 oscillator, 343; 554
 switch, 155

AP 3302 PART 3

Turnstile junction, 301
Twinline tracking, Doppler, 491–494
Twists, waveguide, 295
TWT (see Travelling wave tubes)

U

UHF aerials, 237–247 (see also Aerials)
UHF radar, 227–235
Ultra audion oscillator, 364
Ultrasonic delay lines, 504
Unijunction timebase generator, 199
Unijunction transistor, 152
 as voltage stabiliser, 393
Unipolar video, 502
Up-converter, parametric, 353

V

Varactor diode, 344; 550
 beam steering, 334
 characteristics, 552
 frequency multiplier, 346
 microwave switch, 345
 parametric amplifier, 348–353; 551
Variable capacitance diode (see Varactor diode)
V band, 33
Velocity gate (CW radar), 457
Velocity gate stealer, 442
Velocity modulation 255
Video amplifier, 419
Video map, 539
Voltage doubler, 385
Voltage regulation, 389–393
Voltage stabiliser, 389

W

Waveguide, 279–288
 boundary conditions, 279
 circular, 286
 circular modes, 287
 E and H fields, 279
 rectangular, 280–286
 'a' dimension, 284
 attenuation, 285
 'b' dimension, 284
 characteristic impedance, 286
 cut-off wavelength, 285
 dimensions, 285
 field pattern, 282
 guide wavelength, 284
 group velocity, 285
 modes, 283

 phase velocity, 285
 propagation in, 281
 transverse electric (TE) wave, 283
 wall currents, 286
 transverse electromagnetic (TEM) wave, 279
Waveguide components, 289–308
 attenuators, 292
 balanced mixer, 306
 bends, 295
 corners, 295
 directional couplers, 296
 dummy load, 293
 ferrite devices, 302; 581–591
 flexible waveguide, 295
 hybrid ring, 299
 hybrid T junction, 299; 624
 irises, 291
 joints, 293
 loop coupling, 289
 magic-T junction, 299
 measurements, 296
 mixer, 306
 power measurement, 298
 rat race, 299
 rotating joint, 294
 slot coupling, 289
 standing wave ratio indicator, 297
 stubs, 290
 T-junctions, 298
 twists, 295
 wavemeter, 253; 297
Waveguide shutters, 315
Wavemeter, 253; 297
Weather effects, 434
Weather radars, 604–607
White noise, 399
Wideband jamming, 440
Window (ECM), 440
Window coupling, waveguide, 252

X

X band, 33; 230
X ray, 355

Y

YIG filters, 590
Y-junction circulator, 584
Yttrium, 591

Z

Zener diode, 390